# LINEAR OPERATORS IN APPLIED ENGINEERING MATHEMATICS
(Lecture Notes)

K. G. Ayappa
Department of Chemical Engineering
Indian Institute of Science
Bangalore

# Contents

# Chapter 1

# Introduction to matrix, differential and integral equations

Matrix, differential and integral equations arise out of models developed to describe various physical situations. A few examples of these equations that arise in the analysis of engineering problems are illustrated in this Chapter. Since the course is mainly concerned with linear systems we will conclude this Chapter with a formal definition of linear operators.

## 1.1 Matrix Equations

Consider the following collection of linear equations, which can be compactly written in matrix vector notation as

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{1.1}$$

where,

$$\mathbf{A}(n \times n) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

We are interested in finding $\mathbf{x}$ given the matrix $\mathbf{A}$ and vector $\mathbf{b}$. The matrix equation represents a collection of linear algebraic equations which arises out of models developed to describe a wide variety of physical situations. These include chemical reactions, staged processes such as distillation and gas absorption, electrical networks and normal mode vibrational analysis of molecules. Matrix equations also arise during numerical solutions of differential equations

using finite difference, finite volume and finite element methods as well as in the numerical solution of integral equations using quadrature methods.

## 1.2 Differential Equations

**Reaction-Diffusion Equation**

Consider the first order reaction, A $\rightarrow$ B, occurring in the inner surface of a cylindrical catalyst pellet of radius $R$ and length $L$ as shown in Fig. 1.1. Performing a mass balance on a

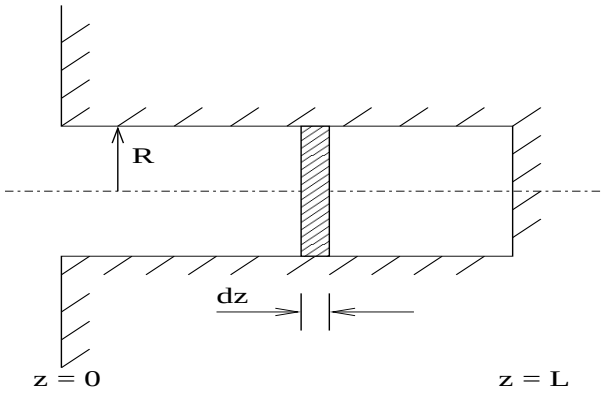



Figure 1.1: Schematic of catalyst pellet of radius $R$ and length $L$

differential element of thickness $\Delta z$ for species A;

$$\pi R^2 \Delta z \frac{\partial C_A}{\partial t} = j_z \pi R^2 - j_{z+\Delta z} \pi R^2 - k_1 C_A 2\pi R \Delta z \tag{1.2}$$

where $j_z$ is the mass flux of the species A and $k_1$ is the first order reaction rate constant. Dividing Eq. 1.2 with $\pi R^2 \Delta z$ and using Ficks law,

$$j_z = -D_{AB} \frac{\partial C_A}{\partial z} \tag{1.3}$$

Eq. 1.2 reduces to

$$\frac{\partial C_A}{\partial t} = D_{AB} \frac{\partial^2 C_A}{\partial z^2} - \alpha_1 C_A \tag{1.4}$$

where $\alpha_1 = 2k_1/R$. Eq. 1.4 is the unsteady state reaction diffusion equation whose solution yields the concentration $C_A(z,t)$. We further note that Eq. 1.4 is a partial differential equation

whose complete formulation requires one initial condition (IC) and two boundary conditions (BCs) to be specified. The initial condition is

$$C_A(z, t = 0) = 0 \tag{1.5}$$

The boundary conditions are

$$C_A(z, t) = C_{A0} \qquad \text{at} \qquad z = 0 \tag{1.6}$$

and

$$\frac{dC_A}{dz} = 0 \qquad \text{at} \qquad z = L \tag{1.7}$$

Eq. 1.7 assumes that the face of the pore at $z = L$ is non-reactive.

*Question: Modify the boundary condition for a reactive pore end at $z = L$.*

Eq. 1.4 is an example of a partial differential equation (PDE) since the dependent variable, $C_A(x, t)$ depends on more than one independent variable $(x, t)$. At steady state, the equation reduces to the following ordinary differential equation (ODE),

$$D_{AB}\frac{d^2C_A}{dz^2} - \alpha_1 C_A = 0 \tag{1.8}$$

Eq. 1.8 along with the boundary conditions at $z = 0$ and $z = L$ constitute what is commonly referred to as a 2 point boundary value problem (BVP) since the boundary conditions at two ends of the pore are required to complete the problem specification.

**Unsteady State Heat Conduction Equation**

Consider a three dimensional solid object, $\Omega$ heated with an internal source $p(x, y, z)$ as shown in Fig. 1.2. The unsteady state heat conduction equation is,

$$\rho C_p \frac{\partial T}{\partial t} = \nabla \cdot k\nabla T + p(x, y, z) \tag{1.9}$$

where $\rho$ is the density, $C_p$ is the specific heat capacity and $k$ the thermal conductivity, which can in general be a function of the spatial co-ordinates. The solution of Eq. 1.9 with appropriate initial and boundary conditions yield the temperature $T(x, y, z, t)$. Eq. 1.9 is an example of a partial differential equation that arises in conduction heat transfer. Unlike the two-point BVP discussed earlier the boundary condition for the heat equation is specified on the entire boundary

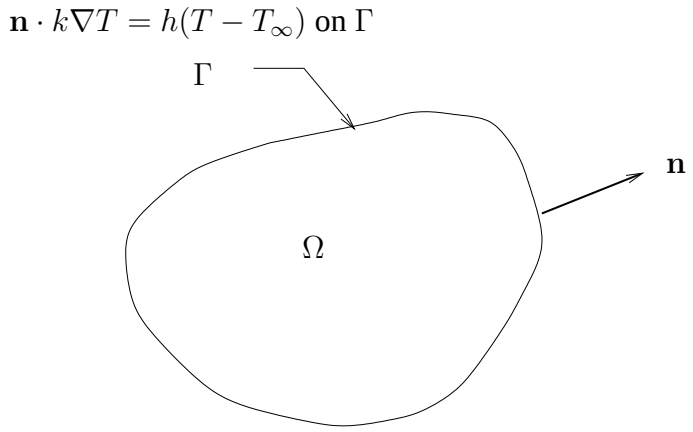



Figure 1.2: Heat conduction in a 3D object. $\Omega$ denotes the domain and $\Gamma$ the surface of the object. At the surface heat is lost by convection. $h$ is the heat transfer coefficient.

$\Gamma$ as illustrated in Fig. 1.2. If the heat transfer coefficient is independent of the spatial co-ordinates Eq. 1.9, reduces to

$$\rho C_p \frac{\partial T}{\partial t} = k\nabla^2 T + p(x, y, z) \tag{1.10}$$

At steady state, in the absence of the source term, $p(x, y, z)$, the heat conduction equation reduces to the Laplace equation,

$$\nabla^2 T = 0 \tag{1.11}$$

Note: The gradient operator in Cartesian co-ordinates is defined as,

$$\nabla T = \mathbf{e}_x \frac{\partial T}{\partial x} + \mathbf{e}_y \frac{\partial T}{\partial y} + \mathbf{e}_z \frac{\partial T}{\partial z} \tag{1.12}$$

and the Laplacian is

$$\nabla^2 T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \tag{1.13}$$

**The Schrödinger Wave Equation**

The wave equation forms the cornerstone of quantum mechanics. It arises in the description of atomic particle positions and quantization of energy levels. The time independent Schrödinger wave equation is

$$\frac{h^2}{8\pi^2 m}\nabla^2 \Psi_n + (E_n - V)\Psi_n = 0 \tag{1.14}$$

where $h$ is Planck's constant, $m$ is the particle mass, $V$ is the potential energy field in which the particle is located. $E_n$ are the corresponding energy levels of the particle and $\Psi_n(\mathbf{r})$ is

the wavefunction which is a function of the spatial co-ordinates. The probability of locating a particle in a volume element $d\mathbf{r}$ is $\Psi(\mathbf{r})\Psi^*(\mathbf{r})\,d\mathbf{r}$, where $\Psi^*(\mathbf{r})$, denotes the complex conjugate of $\Psi(\mathbf{r})$.

## 1.3 Integral equations

Many physical situations naturally give rise to integral equations. Integral equations can sometimes be derived from differential equations. Integrals equations can be broadly classified into Volterra and Fredholm type equations. The Volterra integral equation of the first kind is,

$$\int_0^x k(x,y)u(y)\,dy = f(x) \tag{1.15}$$

where $k(x,y)$ is the kernel of the operator, $f(x)$ is usually some known function and $u(y)$ the solution we seek lies in the integrand. The kernel of the operator is related to the physics of the problem that results in the integral equation. A characteristic feature of the Volterra equation is that the upper limit of the integral is not a fixed quantity.

The Fredholm integral equation of the first kind is,

$$\int_a^b k(x,y)u(y)\,dy = f(x) \tag{1.16}$$

The main difference between the Volterra equation, Eq. 1.15 and the Fredholm integral equation, Eq. 1.16 is that the limits of the integral in the Fredholm equation are fixed. Broadly speaking the Volterra type integral equations are related to initial value problems (IVPs) giving rise to the variable upper limit in the integral of Eq. 1.15 and the Fredholm type equations are related to boundary value problems (BVPs). The final point to note is that integral equations do not require the specification of any additional initial and/or boundary conditions. These conditions are built into the integral equations themselves.

One final classification that is important is that both the integral equations given above were referred to as first kind equations. The general form of a Volterra integral equations of the second kind is

$$\int_0^x k(x,y)u(y)\,dy + \alpha u(x) = f(x) \tag{1.17}$$

where $\alpha$ is an arbitrary scalar quantity, and the second kind Fredholm equation is,

$$\int_a^b k(x,y)u(y)\,dy + \alpha u(x) = f(x) \tag{1.18}$$

In the second kind equations the unknown $u(x)$ appears both inside and outside the integrand.

**Example:** The relationship between an IVP and Volterra operators can be illustrated with a simple example. Consider the following IVP,

$$\frac{du}{dt} + \alpha u = 0 \qquad u(t=0) = u_0 \tag{1.19}$$

where $\alpha$ and $\beta$ are constants. Integrating Eq. 1.19 and using the IC, Eq. 1.19 can be transformed into the following Volterra integral equation of the second kind,

$$\int_0^t \alpha u(y)\, dy + u(t) = u_0 \tag{1.20}$$

The integral equation has a simple kernel and satisfies the initial conditions of Eq. 1.19. Further the solution to Eq. 1.19 is, $u(t) = u_0 \exp(-\alpha t)$. Naturally this is also a solution to its equivalent integral equation, Eq. 1.20.

## 1.4 Linear Operators

Our primary concern as engineers is to obtain solutions to the different classes of equations presented above. Presented with an equation the natural question one poses is whether the equation is solvable or not. This is the problem of existence. If the problem is not solvable, one has to revisit the model assumptions and the underlying physical processes that govern the equation. If the problem is solvable, we inquire if the solution is unique. These questions of existence and uniqueness form a general theme in this book. Most of us are familiar with these ideas with the solution of linear equations of the form given in Eq. 5.59. Can we now generalize these ideas to a more general class of equations? For example under what conditions can one examine the existence and uniqueness conditions for Eq. 1.4 or Eq. 1.15? A unifying theory that integrates all the above questions about existence and uniqueness is the theory of linear operators, which will form an underlying and unifying theme for the course. The theory is general and as long as the operator is linear it will be applicable. A class of linear operators which will require a somewhat more specialized treatment are partial differential equations. Before we proceed further, let us formally define a linear operator.

Consider the generic equation

$$Lu = f \tag{1.21}$$

where $L$ is the operator, $u$ is the solution we seek and $f$ is usually specfied as part of the problem definition. In the matrix equation $L \equiv \mathbf{A}$, $u \equiv \mathbf{x}$ and $f \equiv \mathbf{b}$ $L$ is said to be a linear operator or linear transformation if it satisfies the following properties

$$\begin{aligned} L(\alpha u) &= \alpha L u \qquad\qquad u \in X \ \forall\, \alpha \\ L(u+v) &= Lu + Lv \qquad u, v \in X \end{aligned}$$

where $X$ is a linear space on which the operator $L$ acts. We will return to a formal definition of linear spaces which contain both vectors and functions later in the text. $L$ can also be looked upon as a mapping of elements in $X$ into itself, $L : X \to X$ and $\alpha$ lies in the associated complex scalar field of the operator[1]. We note that both properties must be satisfied for the operator to be linear. Further one property does not imply the other. If either of the above two properties are not satisfied the operator is said to be nonlinear. Both the above requirements of a linear operator can be integrated into a single property,

$$L(\alpha u + \beta v) = \alpha L u + \beta L v \qquad u, v \in X \ \forall\, \alpha, \beta \tag{1.22}$$

**Example 1:** The identity operator maps elements in $X$ into itself.

$$\mathbf{I}u = u \qquad \forall u \in X \tag{1.23}$$

To show that **I** is linear, we note that

$$\begin{aligned} \mathbf{I}(\alpha u + \beta v) &= \alpha \mathbf{I} u + \beta \mathbf{I} v \\ &= \alpha u + \beta v \end{aligned}$$

**Example 2:** The $(n \times n)$ matrix $\mathbf{A}$ is a linear transformation since

$$\begin{aligned} \mathbf{A}(\alpha \mathbf{u} + \beta \mathbf{v}) &= \sum_{j=1}^{n} a_{ij}(\alpha u_j + \beta v_j) \qquad i = 1 \ldots n \\ &= \alpha \sum_{j=1}^{n} a_{ij} u_j + \beta \sum_{j=1}^{n} a_{ij} v_j \qquad i = 1 \ldots n \\ &= \alpha \mathbf{A}\mathbf{u} + \beta \mathbf{A}\mathbf{v} \end{aligned}$$

**Example 3:** The reaction diffusion equation, where the operator

$$Lu = \frac{\partial u}{\partial t} - D_{AB} \frac{\partial^2 u}{\partial z^2} + \alpha_1 u \tag{1.24}$$

$$\begin{aligned} L(\alpha u + \beta v) &= \frac{\partial(\alpha u + \beta v)}{\partial t} - D_{AB}\frac{\partial^2(\alpha u + \beta v)}{\partial z^2} + \alpha_1(\alpha u + \beta v) \\ &= \alpha(\frac{\partial u}{\partial t} - D_{AB}\frac{\partial^2 u}{\partial z^2} + \alpha_1 u) + \beta(\frac{\partial v}{\partial t} - D_{AB}\frac{\partial^2 v}{\partial z^2} + \alpha_1 v) \\ &= \alpha L u + \beta L v \end{aligned}$$

Hence the differential operator is linear. Since the operator involves both the differential equation and its associated initial and boundary conditions, the IC and BCs must also satisfy the linearity property for the differential equation to be classified as linear. This can easily be verified for the reaction diffusion equation diffusion, Eq. 1.4.

The linearity property of differential operators, has one important consequence from the viewpoint of obtaining solutions. It simply means that if $u$ and $v$ are solutions of the differential equation, then $w = \alpha u + \beta v$ is also a valid solution. This idea, technically called the *principle of superposition*, is used widely in the solution of of both ordinary and partial differential equations. A familiar example is the solution of the linear differential equation

$$\frac{d^2u}{dx^2} - m^2 u = 0$$

The solution to the above equation, $u(x) = c_1 e^{mx} + c_2 e^{-mx}$. Since the differential equation is linear, not only are $u_1 = e^{mx}$ and $u_2 = e^{-mx}$ independent solutions, $u = c_1 u_1 + c_2 u_2$ is also a valid solution. The constants $c_1$ and $c_2$ are obtained by using the appropriate boundary or initial conditions. The principle of superposion has the following consequence for linear operators. If

$$u = \sum_{i=1}^{n} c_i u_i$$

then,

$$Lu = L\sum_{i=1}^{n} c_i u_i = \sum_{i=1}^{n} c_i L u_i$$

**Example 4:** Volterra Integral equation (Eq. 1.15).

$$\begin{aligned} L(\alpha u + \beta v) &= \int_0^x k(x,y)[\alpha u(y) + \beta v(y)]\, dy \\ &= \int_0^x k(x,y)\alpha u(y) dy + \int_0^x k(x,y)\beta v(y)\, dy \\ &= \alpha \int_0^x k(x,y)u(y)\, dy + \beta \int_0^x k(x,y)v(y)\, dy \\ &= \alpha Lu + \beta Lv \end{aligned}$$

**Example 5:** To show that if an operator satisfies the property $L(u+v) = Lu + Lv$, it need not satisfy $L(\alpha u) = \alpha Lu$. Consider $L$ to be the operation of complex conjugation. If $z$ and $w$ are two complex numbers then $L(z+w) = Lz + Lw$. Clearly $L(\alpha z) \neq \alpha Lz$ if $\alpha$ is a complex scalar.

**Example 6:** To show that if an operator satisfies the property $L(u+v) = Lu + Lv$, it need not satisfy $L(\alpha u) = \alpha Lu$. Consider the operation of mapping the components [2], $x_1$, $x_2$ of a 2d vector into a point;

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} x_1 + x_2 & x_1 x_2 > 0 \\ \\ 0 & \text{otherwise} \end{cases}$$

Assuming a real field of scalars $\alpha$, and vectors $u = (1,-1)$ and $v = (1,1)$, $L(u+v) = 0$ whereas, $Lu + Lv = 1$.

Before we conclude, we briefly discuss the classifications of equations into homogeneous and inhomogeneous. In the generic equation, Eq. 1.21 the equation is homogeneous if the right hand side, $f$ = 0. Hence $\mathbf{Ax} = \mathbf{0}$ is an example of a homogeneous set of linear equations. In the case of differential equations $f$ represents the term containing only the independent variables and $L$ represents the operator acting on the dependent variable. Hence the reaction-diffusion equation, Eq. 1.4 and the Schrödinger wave equation, Eq. 1.14 are examples of homogeneous differential equations. The unsteady state heat equation, Eq. 1.9 is inhomogeneous due to the presence of the source term $p(x,y,z,t)$ and the integral equations given in Eqs. 1.15 and Eq. 1.16 are both inhomogeneous integral equations. In the case of differential equations the ICs and BCs can also be classified as homogeneous and inhomogeneous in a similar manner. This classification is important, as we will soon see that determining the solu-

tions to the homogeneous problem forms the first step in studying the existence and uniqueness conditions for the inhomogeneous problem.

## 1.5 Summary

In this Chapter we have introduced some simple examples of various kinds of equations that commonly arise in engineering and sciences. Starting with matrix equations, the basic difference between ODEs and PDEs should be clear from the examples presented above. ODEs can further be classified into IVP's where all the conditions are specified as an initial condition at time $t = 0$ and BVP's where the differential equation is accompanied by boundary conditions. The classification presented in here is preliminary. PDEs can be further classified into various categories and these will be discussed later in the text. Some examples and a preliminary classification of integral equations was also introduced. Integral equations are more specialized and do not arise as often in the description of physical problems as do differential equations. Sometimes integral equations have an equivalent representation as a differential equation as we encountered with the IVP problem. In many cases this equivalence is not feasible and the integral equation has to be solved directly. Linearity is an important concept and its consequence, the principle of superposition, used routinely for solving linear operators should be recognized. Finally the notion of homogeneous and inhomogeneous equations should be firmly understood.

# Bibliography

[1] Ramkrishna D and N. R. Amundson. *Linear Operator Methods in Chemical Engineering*. Prentice Hall, 1985.

[2] Naylor A. W and G. R. Sell. *Linear Operator Theory in Engineering and Science*. Springer-Verlag New York, 1982.

**PROBLEMS**

1. If $L$ is a linear operator show that $L^n$ is also a linear operator. Note that $L^2 = LL$ and so on. Use the method of mathematical induction for your proof.

2. Using the properties of a linear operator, $L(u+v) = Lu + Lv$ and $L(\alpha u) = \alpha L(u)$, identify which of the following operators are linear.

   (a)
   $$Lu = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + \left(e^{-x} + x^2\right)\frac{\mathrm{d}u}{\mathrm{d}x} + xu.$$

   (b)
   $$Lc = \frac{\partial c}{\partial t} - \frac{\partial}{\partial x}\left(D\left(c\right)\frac{\partial c}{\partial x}\right) - kc$$

   where $D(c)$ is a concentration dependent diffusion coefficient and $k$ is the reaction rate constant. Rework this problem with $D$ as a function only of $x$.

   (c)
   $$Lu \equiv \int_0^x \frac{u(y)}{\sqrt{\alpha x - \beta y}} dy = f(x)$$

Differentiate the above Volterra integral equation using the rules for differentiation under an integral sign. Is the resulting operator still linear? The process of differentiation converts the first kind Volterra integral equation to that of the second kind, where the unkown $u$ appears both inside the integral as well as outside. Note: $\alpha$ and $\beta$ are arbitrary scalars.

3. Using the properties of linear operators, determine which of the following operators are linear

   (a) The divergence operator,
   $$Lu \equiv \nabla \cdot \alpha(x, y, z) \nabla u$$

   (b) The curl operator,
   $$Lu \equiv \nabla \times \mathbf{u}$$

   (c) The Fredholm integral operator
   $$Lu \equiv \int_a^b e^{x-y}\, u(x) dx + u^2(y)$$

4. In the following equations of the generic form $Lu = f$, identify the operator and determine if the operator is linear or not.

   (a) Heat equation with spatially dependent thermal conductivity
   $$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\sin x\, e^{-x}\right) \frac{\partial u}{\partial x} = f(x)$$

   (b) The $n$th order ordinary differential equation
   $$a_0 \frac{\partial^n u}{\partial x^n} + a_1 \frac{\partial^{n-1} u}{\partial x^{n-1}} + \cdots + a_n = f(x)$$

   (c) The 3D wave equation:
   $$\frac{\partial^2 u}{\partial t^2} = c^2 \nabla^2 u$$

   (d) Integral equation:
   $$\int_0^x e^{(x-y)} dy\, e^{u(y)} + u(x) = f(x)$$

(e) The Integro-differential equation,

$$\frac{du}{dt} = \int_0^t e^{\frac{t-t'}{\tau}}\, u(t')\, dt'$$

(f) The Korteweg -de Vries (KDV) equation used in the study of water waves

$$\frac{\partial u}{\partial t} + cu\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0$$

(g) Which of the above equations given in parts (a) - (f) are homogeneous.

5. Check the following transforms for linearity

(a) The Laplace transform

$$f(s) \equiv L[f(t)] = \int_0^\infty e^{st} f(t)\, dt$$

(b) The Fourier transform

$$f(\omega) \equiv L[f(t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} f(t)\, dt$$

6. Using the following dimensionless variables,

$$u = C_A(z)/C_{A0}, x = z/L$$

the dimensionless form of the steady state differential equation (Eq. 1.8) is,

$$\frac{d^2u}{dx^2} - \phi^2 u = 0 \qquad 0 < x < 1 \tag{1.25}$$

where $\phi^2 = 2k_1L^2/D_{AB}R$, and $L$ is the pore length and $R$ is the radius of the pore. Obtain analytical solutions for both the dead end pore and the reactive end pore. Qualitatively sketch your solution for the dimensionless concentration $u$ for small and large values of the parameter $\phi^2$. Physically interpret these two conditions.

# Chapter 2

# Properties of Matrices

In this section we review some basic properties of matrices. Let $\mathbf{A}$ be a $m \times n$ matrix,

$$\mathbf{A}(m \times n) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

where $m$ is the number of rows and $n$ the number of columns. $a_{ij}$ will the $i, j$ element in the matrix. A column vector $\mathbf{x}$ is an $n \times 1$ matrix,

$$\mathbf{x}(n \times 1) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Matrices arise frequently in engineering applications and can assume a variety of forms, some of which are illustrated below. Many of these forms arise during numerical solution of differential equations and recognizing the form of the matrix is important while choosing the appropriate solution technique.

## 2.1 Equality of matrices

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be equal to each other if $a_{ij} = b_{ij}$. Only matrices of similar order can be considered to be equal.

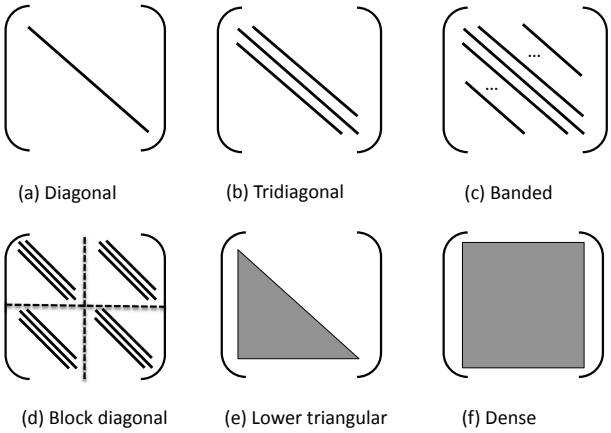



Figure 2.1: Various classifications of commonly occurring matrices. The solid lines and filled regions represent non zero elements. Sparse matrices (not shown) contain a sparse distribution of non-zero elements in the matrix.

## 2.2 Addition of matrices

Matrices are compatible for addition only if the corresponding numbers of rows and columns are similar. Matrix addition is both associative and commutative

$$1a \qquad (\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C}) \qquad \text{Associative}$$

$$1b \qquad \mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \qquad \text{Commutative}$$

## 2.3 Scalar multiplication

When a matrix is multiplied by a scalar $\alpha$ all the elements of the matrix are multiplied $\alpha$.

$$\alpha\mathbf{A}(m\times n)=\begin{pmatrix}\alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{1n}\\ \alpha a_{21} & \alpha a_{22} & \dots & \alpha a_{2n}\\ \vdots & \vdots & & \vdots\\ \alpha a_{m1} & \alpha a_{m2} & \dots & \alpha a_{mn}\end{pmatrix}$$

## 2.4 Multiplication of Matrices

Two matrices $\mathbf{A}(m\times n)$ and $\mathbf{B}(n\times p)$ are compatible for multiplication if the number of columns of $\mathbf{A}$ are similar to the number of rows of $\mathbf{B}$. Matrix multiplication satisfies the following

properties

$$3a \qquad \mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$$

$$3b \qquad (\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$$

$$3c \qquad (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

$$3d \qquad \text{In general} \quad \mathbf{AB} \neq \mathbf{BA}$$

## 2.5 Transpose of a matrix

The transpose of a matrix $\mathbf{A}$ is obtained by interchanging its rows and columns. The transpose is denoted by $\mathbf{A}^T$. The operation of a transpose satisfies the following properties.

$$4a \qquad (\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$

$$4b \qquad (\mathbf{A}^T)^T = \mathbf{A}$$

$$4c \qquad (\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$$

## 2.6 Trace of a matrix

The sum of the diagonal elements of a square matrix is known as the trace. The trace of an $(n \times n)$ square matrix,

$$\text{Trace}\mathbf{A} = \sum_{i=1}^{n} a_{ii}$$

A number of the properties of matrices listed above can be proved using index algebra. We illustrate these manipulations with some examples which the reader should get acquainted with.

**Example**: Matrix vector multiplication

$$\mathbf{Ax} = \mathbf{b} \rightarrow \sum_{j=1}^{n} a_{ij}\, x_j = b_i \qquad i = 1 \ldots m$$

where $\mathbf{A}$ is an $(m \times n)$ matrix and $\mathbf{x}$ is $(n \times 1)$ and $\mathbf{b}$ has dimenions $(m \times 1)$.

**Example**: Matrix multiplication

$$\mathbf{A}(m \times p)\mathbf{B}(p \times n) = \mathbf{C}(m \times n) \rightarrow \sum_{k=1}^{p} a_{ik}\, b_{kj} = c_{ij} \quad i = 1 \ldots m, \quad j = 1 \ldots n$$

**Example:** To show that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$. Let $c_{ij}^T$ be the element of $(\mathbf{AB})^T$

$$\begin{aligned} c_{ij} &= \sum_{k=1}^{n} a_{ik}b_{kj} \\ c_{ij}^T &= \sum_{k=1}^{n} a_{jk}b_{ki} \end{aligned}$$

Let $d_{ik}$ be the element of $\mathbf{B}^T\mathbf{A}^T$. We need to show that $d_{ij} = c_{ij}^T$

$$\begin{aligned} d_{ik} &= \sum_{j=1}^{n} b_{ji}a_{kj} \\ d_{ij} &= \sum_{k=1}^{n} b_{ki}a_{jk} \\ &= \sum_{k=1}^{n} a_{jk}\, b_{ki} = c_{ij}^T \end{aligned}$$

The second line in the algebra above is obtained by interchanging the index $k$ with $j$. This example illustrate manipulations with indices that the reader should be acquainted with.

**Example:** To show that, Trace $(\mathbf{AB})$ = Trace $(\mathbf{BA})$

$$\begin{aligned} \text{Trace}\,(\mathbf{AB}) &= \sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}b_{ki} \\ \text{Trace}\,(\mathbf{BA}) &= \sum_{i=1}^{n}\sum_{k=1}^{m} b_{ik}a_{ki} \\ &= \sum_{i=1}^{m}\sum_{k=1}^{n} b_{ki}a_{ik} \\ &= \text{Trace}\,(\mathbf{AB}) \end{aligned}$$

We have assumed that $\mathbf{A}$ is an $(m \times n)$ matrix and $\mathbf{B}$ is an $(n \times m)$ matrix. In the second line of the above equation, the indices have been exchanged and the upper limits in the summation have been consistently altered.

## 2.7 Symmetric and Hermitian Matrices

The matrix **A**, is said to be symmetric if

$$\mathbf{A} = \mathbf{A}^T. \tag{2.1}$$

We note that the above notion of symmetry is restricted to real matrices. If the matrix has complex elements then we define $\mathbf{A}^*$ as the matrix obtained by taking the complex conjugate of $\mathbf{A}^T$. Hence

$$\mathbf{A}^* = \overline{\mathbf{A}}^T \equiv \overline{\mathbf{A}^T}$$

Note that the operation of taking the transpose and complex conjugation commute.

A matrix is said to be Hermitian if

$$\mathbf{A} = \mathbf{A}^* \tag{2.2}$$

The above definition includes real matrices as well. In the case of real matrices Eq. 2.2 is equivalent to Eq. 2.1.

**Example**:

$$\mathbf{A} = \begin{pmatrix} 1 & i \\ i & 2 \end{pmatrix} \qquad \mathbf{A}^* = \begin{pmatrix} 1 & -i \\ -i & 2 \end{pmatrix}$$

**Example**:

$$\mathbf{A} = \begin{pmatrix} i & 1-i \\ 1+i & 0 \end{pmatrix} \qquad \mathbf{A}^* = \begin{pmatrix} -i & 1-i \\ 1+i & 0 \end{pmatrix}$$

**Example**:

$$\mathbf{A} = \begin{pmatrix} 1 & 1-i \\ 1+i & 0 \end{pmatrix} \qquad \mathbf{A}^* = \begin{pmatrix} 1 & 1-i \\ 1+i & 0 \end{pmatrix}$$

**Example**:

$$\mathbf{A} = \begin{pmatrix} i & i \\ i & i \end{pmatrix} \qquad \mathbf{A}^* = -\begin{pmatrix} i & i \\ i & i \end{pmatrix}$$

Only the matrix in the 3rd example is Hermitian. Clearly a matrix with complex elements on the diagonal cannot be Hermitian. The last example is an example of a skew Hermitian matrix where $\mathbf{A}$ = -$\mathbf{A}^*$

**Example:** To show that the product of two symmetric matrices need not be symmetric. Let $\mathbf{A}$ and $\mathbf{B}$ be two symmetric matrices.

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T = \mathbf{BA} \neq \mathbf{AB}$$

This is an example of a proof which did not involve the use of indices.

## 2.8 Inverse

The inverse of $\mathbf{A}$ denoted by $\mathbf{A}^{-1}$ is such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

where the identity matrix **I** is a diagonal matrix with 1's on the diagonal.

**Example**

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

where $|\mathbf{A}| = a_{11}a_{22} - a_{21}a_{12}$ is the determinant of the matrix $\mathbf{A}$. We can generalize the definition of the inverse by using the adjoint or adjugate of a matrix.

## 2.9 Determinants, Cofactors and Adjoints

The minor $|M_{ij}|$ of an element $a_{ij}$ in the matrix $\mathbf{A}$ is the determinant of an $(n-1) \times (n-1)$ matrix formed by omitting the $i^{th}$ row and the $j^{th}$ column. The cofactor of the element $a_{ij}$,

$$A_{ij} = (-1)^{i+j}|M_{ij}|$$

The determinant of an $n \times n$ matrix expressed as an expansion in terms of the cofactors is

$$\begin{aligned} |\mathbf{A}| &= \sum_{j=1}^{n} a_{1j}A_{1j} \\ &= \sum_{j=1}^{n} a_{1j}(-1)^{1+j}|M_{1j}| \end{aligned}$$

The adjoint of a matrix, denoted as adj **A**, is the transpose of the cofactor matrix whose elements are made up of $A_{ij}$ (Eq. 2.9). This definition should not be confused with the adjoint operator whose definition we will encounter later in the text. The inverse of a matrix can be expressed using the definition of the adjoint by noting that

$$\begin{aligned} \mathbf{A}\,\text{adj}\mathbf{A} &= |\mathbf{A}|\mathbf{I} \\ (\text{adj}\mathbf{A})\,\mathbf{A} &= |\mathbf{A}|\mathbf{I} \end{aligned}$$

which implies that

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\text{adj}\mathbf{A}$$

**Example**

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \\ 1 & -1 & 1 \end{pmatrix}, \qquad \text{adj}\mathbf{A} = \begin{pmatrix} 5 & -1 & 7 \\ 2 & 2 & -2 \\ -3 & 3 & 3 \end{pmatrix}, \qquad \mathbf{A}^{-1} = \frac{1}{12}\text{adj}\mathbf{A}$$

## 2.10 Echelon forms, rank and determinants

The echelon form for any matrix $\mathbf{A}$ is such that the number of zeroes preceding the first non-zero element in every row increases row by row (starting from the 1st row). The echelon forms can be obtained by performing elementary row operations on the matrix. Some examples of row reduce echelon forms are given below,

$$\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 & 5 & 2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.3}$$

Clearly if the number of zeroes preceding the first non-zero element is the same in the $k$ and $k+1$ rows, the first non-zero element in the $k+1$ row can be made zero by elementary row operations. Once the echelon forms are obtained it is easy to deduce the rank of the matrix. The rank $r$ of a matrix $\mathbf{A}$ is the number of rows containing non-zero elements in the row reduced echelon form. Using the above definition, the rank of the matrices given above are 2, 1 and 3 respectively. From the definition of the rank it is easy to see that $r \leq min(m, n)$, $r \neq 0$ (unless all the elements in the matrix are identically zero). A similar definition of the rank can be generated using column operations. The row rank is equal to the column rank or equivalently the maximum number of linearly independent rows is equal to the maximum number of linearly independent columns in a matrix. In subsequent use of the rank, we will use the definition based on the row rank as this will provide a convenient method for obtaining solutions of linear systems of equations. An additional definition of the rank is based on the determinant. The rank is the order of the largest non-zero determinant in the matrix. If the rank of a matrix is $k$, then there is at least one determinant of order $k$ that is nonzero. All determinants of order $k+1$ must vanish.

Some examples of obtaining the ranks of matrices using the echelon forms are given below,

Example: $(3 \times 3)$ matrix, with $r = 3$.

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \\ 1 & 0 & 1 \end{pmatrix} \quad \begin{matrix} R_3 - R_1 \\ \longrightarrow \end{matrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \\ 0 & -2 & 2 \end{pmatrix} \quad \begin{matrix} 3R_3 + 2R_2 \\ \longrightarrow \end{matrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \\ 0 & 0 & 10 \end{pmatrix} \tag{2.4}$$

Example: $(3 \times 3)$ matrix, with $r = 2$.

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 2 \\ 1 & 1 & 3 \end{pmatrix} \quad \begin{matrix} R_2 - 2R_1 \\ \longrightarrow \\ R_3 - R_1 \end{matrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 4 \\ 0 & -1 & 4 \end{pmatrix} \quad \begin{matrix} R_3 - R_2 \\ \longrightarrow \end{matrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 4 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.5}$$

The rank of the above matrix can be obtained by column operations,

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & 2 \\ 1 & 1 & 3 \end{pmatrix} \quad \begin{matrix} C_2 - C_1 \\ \longrightarrow \\ C_3 - 3C_1 \end{matrix} \quad \begin{pmatrix} 1 & 1 & -4 \\ 2 & 1 & -4 \\ 1 & 0 & 0 \end{pmatrix} \quad \begin{matrix} C_3 + 4C_2 \\ \longrightarrow \end{matrix} \quad \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{2.6}$$

The above example illustrates that the rank can be determined by either the row or column reduced echelon forms. This is a consequence of the property that the order of the smallest non-zero determinant of a matrix is unchanged by elementary row or column operations as carried out above (show this). Some elementary properties of determinants can be understood from the above examples. For a square matrix $(n \times n)$ the determinant is non-zero if and only if $r = n$. Adding a multiple of one row to another leaves the determinant unchanged. Thus in Eq. 2.4 above, the determinant of the matrix is 10, row operations $R_3 - R_1$ leaves the determinant unchanged, however the last row operation $3R_3 + 2R_2$ changes the determinant to 30, since $R_3$ is multiplied by 3. The last property can easily be proved with the help of cofactor expansions. To show that row operations, $\alpha R_1 + \beta R_2$ leaves the determinant of an $n \times n$ matrix multiplied by $\alpha$,

$$\begin{aligned} |\mathbf{A}| &= \sum_{j=1}^{n} (\alpha a_{1j} + \beta a_{2j}) A_{1j} \\ &= \sum_{j=1}^{n} \alpha a_{1j} A_{1j} + \sum_{j=1}^{n} \beta a_{2j} A_{1j} \\ &= \alpha \sum_{j=1}^{n} a_{1j} A_{1j} \\ &= \alpha |\mathbf{A}| \end{aligned}$$

Since the first row of the matrix is a multiple of the second row (rank = 0), the second term in the second line above is identically zero. If $\alpha = 1$ then the determinant is unchanged.

**PROBLEMS**

1. Consider the matrix
$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 3 & 5 \\ 2 & 1 & -1 \end{pmatrix}$$
   (a) Find $\mathbf{A^T}$, $\mathbf{A}^{-1}$, $\mathbf{A}^2$, det$(\mathbf{A})$ and det$(\mathbf{A}^5)$. Use the adjoints to find the inverse.
   (b) Find the solution to $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{b} = (-1, 1, 3)^T$

2. Consider the matrix
$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 3 & 2 \\ 1 & 1 & -4 \end{pmatrix}$$
   Find the solutions to $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{b} = (1, 2, 0)^T$

3. A skew symmetric matrix is such that
$$\mathbf{A^T} = -\mathbf{A}$$
   (a) Show that a skew symmetric matrix is square.
   (b) What are the diagonal elements of a skew symmetric matrix ?
   (c) If $\mathbf{A}$ is an $(n \times n)$ matrix then show that $(\mathbf{A} - \mathbf{A^T})$ is skew symmetric.
   (d) Show that any square matrix can be decomposed into a sum of symmetric and skew symmetric matrices.
   (e) Show that a Hermitian matrix can be written as the sum of a real symmetric matrix and an imaginary skew symmetric matrix. Check this property with a suitable example.

4. Show that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

5. Using the definition of cofactors and adjoints show that
$$\mathbf{A}(\text{adj}\mathbf{A}) = (\text{adj}\mathbf{A})\mathbf{A} = |\mathbf{A}|\mathbf{I}.$$

6. If $\mathbf{A}$ and $\mathbf{B}$ are two noncommuting Hermitian matrices such that

$$\mathbf{AB} - \mathbf{BA} = i\mathbf{C},$$

prove that $\mathbf{C}$ is Hermitian.

7. The sum of the diagonal elements in a square matrix is known as the trace. Show that

$$\text{trace}(\mathbf{AB} - \mathbf{BA}) = 0$$

8. If $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices, show that $(\mathbf{AB} + \mathbf{BA})$ and $i(\mathbf{AB} - \mathbf{BA})$ are also Hermitian.

9. If $\mathbf{C}$ is non-Hermitian, show that $\mathbf{C} + \mathbf{C}^*$ and $i(\mathbf{C} - \mathbf{C}^*)$ are Hermitian.

10. A real matrix is said to be orthogonal if $\mathbf{A}^{-1} = \mathbf{A}^T$. Show that the product of two orthogonal matrices is orthogonal. Further, show that $\det(\mathbf{A}) = \pm 1$. Note: If $\mathbf{A}$ is complex and $\mathbf{A}^{-1} = \mathbf{A}^*$ then $\mathbf{A}$ is said to be unitary.

11. Orthogonal matrices arise in co-ordinate transformations. Consider a point $(x, y)$ in the $X - Y$ plane. If the $X - Y$ plane is rotated counter-clockwise by an angle $\phi$ then the point $(x, y)$ is transformed to the point $(x', y')$ in the $X' - Y'$ co-ordinate system. The rotation operation can be represented by a matrix equation

$$\mathbf{Ax} = \mathbf{x}'$$

or

$$\begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

In 3 dimensions, rotation about the $z$ axis by an angle $\phi$ is represented by

$$\mathbf{B} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Verify that $\mathbf{A}$ and $\mathbf{B}$ are orthogonal matrices.

# Chapter 3

# Vector or Linear Spaces

The vector or linear space is the simplest of the abstract spaces that we will encounter. A vector space $X$ is a collection of vectors that can be combined by addition and each vector can be multiplied by a scalar. The elements of a vector space satisfy the following axioms. If $u, v, w \in X$ and $\alpha$ and $\beta$ lie in the associated field of scalars, the elements in the vector space satisfy the following axioms,

**1. Linearity:**

$$(1a) \qquad u + v = v + u$$

$$(1b) \qquad u + (v + w) = (u + v) + w$$

$$(1c) \qquad \text{There } \exists \text{ a unique vector } 0 \text{ such that } u + 0 = u \quad \forall\, u \in X$$

$$(1d) \qquad u + (-u) = 0$$

**2. Multiplication by a scalar:**

$$(2a) \qquad \alpha(\beta u) = \alpha\beta u$$

$$(2b) \qquad (\alpha + \beta)u = \alpha u + \beta u$$

$$(2c) \qquad \alpha(u + v) = \alpha u + \alpha v$$

In order to show that elements in a set $X$ constitute a vector space, the elements must conform to all the properties of the linear space listed above. The properties of a linear space simply allow for vector addition and multiplication of the elements by a scalar. Examples of vectors spaces

are the $n$-dimensional vector space which consists of vectors with $n$ real elements, also referred to as $\mathbf{R}^n$. Alternately the elements that constitute the vector can be complex. This is known as the space $\mathbf{C}^n$. Functions can also make up a linear space. Hence the set of all continuous functions on the interval [a,b] make up a vector space, called $\mathbf{C}[a,b]$. The reader should ensure that these examples satisfy the properties of the linear space

## 3.1 Linear Independence, Basis and Dimension

The notion of linear independence and dependence are important and desirable properties for a collection of vectors. The ideas developed in this section are important while obtaining solutions to linear equations and lay a general framework for obtaining solutions to various classes of operators. A collection of vectors $\mathbf{u}_1, \mathbf{u}_2 \ldots \mathbf{u}_k$ are said to be linearly independent if the only solution to

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 \ldots \alpha_n \mathbf{u}_n = 0 \tag{3.1}$$

is the trivial solution i.e. $\alpha_i = 0$ for $i = 1 \ldots n$. Eq. 3.1 represents a linear combination of vectors. If there exists some values of $\alpha_i$, not all zero, such that Eq. 3.1 is satisfied then the set of vectors are linearly dependent. In other words for the set to be linearly dependent, non trivial solutions exist for Eq. 3.1. We illustrate the notion of linear independence by relating them to solutions of homogeneous linear equations. If $\mathbf{u_i}$ consists of a collection of vectors in $\mathbf{R}^n$ then ,

$$\mathbf{u_1} = \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{n1} \end{pmatrix} \quad \mathbf{u_2} = \begin{pmatrix} u_{12} \\ u_{22} \\ \vdots \\ u_{n2} \end{pmatrix} \quad \ldots \mathbf{u_n} = \begin{pmatrix} u_{1n} \\ u_{2n} \\ \vdots \\ u_{nn} \end{pmatrix}$$

where $u_{ij}$ is the $i$th element of the vector $\mathbf{u}_j$ then Eq. 3.1 can be recast as a collection of homogeneous linear equations which can be represented as

$$\mathbf{A}\alpha = 0 \tag{3.2}$$

where

$$\mathbf{A}(n \times n) = \begin{pmatrix} u_{11} & u_{12} & \ldots & u_{1n} \\ u_{21} & u_{22} & \ldots & u_{2n} \\ \vdots & \vdots & & \vdots \\ u_{n1} & u_{n2} & \ldots & u_{nn} \end{pmatrix} \quad \text{and} \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

Hence in order to examine whether the set of vectors given in Eq. 3.1 is linearly independent we can equivalently seek solutions of the set of linear algebraic equations Eq. 3.2. If Eq. 3.2 has nontrivial solutions ($\alpha_i \neq 0$ for any $i$) then the set of vectors is linearly dependent. If the only solution is the trivial solution ($\alpha_i = 0$ for all $i$) then the set is linearly independent. We illustrate this with some examples

**Example 1**: Consider the set of vectors

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Recasting them into a set of algebraic equations of the form Eq. 3.2,

$$\mathbf{A}\alpha \quad = \begin{pmatrix} 1 & 1 & 0 \\ 2 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \quad = 0$$

Using row operations, it can be shown that the solution to the above equation is only the trivial solution. The determinant of $\mathbf{A}$ is non zero since the rank = 3. Hence the set of vectors are linearly independent.

**Example 2**: Consider the set of vectors

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Recasting them into a set of algebraic equations of the form Eq. 3.2, it can be shown that the non trivial solution is,

$$\alpha = c \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

where $c$ is an arbitrary constant. Hence the set of vectors are linearly dependent. From the last example we can see that if any two vectors in a set are linearly dependent then the entire set is linearly dependent. We can generalize this observation.

**Theorem**: If a subset of vectors in a set of vectors are linearly dependent then the entire set is linearly dependent.

**Proof**: Consider a set of $n$ vectors where the first $m$ vectors are linearly dependent.

$$\sum_{i=1}^{m} \alpha_i u_i + \sum_{i=m+1}^{n} \alpha_i u_i = 0 \tag{3.3}$$

Since the first sum containing vectors from $i = 1$ to $m$ forms a linearly dependent set, this implies that there are values of $\alpha_i$ for which,

$$\sum_{i=1}^{m} \alpha_i u_i = 0$$

Further since the second sum containing terms from $i = m + 1$ to $n$ are linearly independent $\alpha_i = 0$ for $i = m + 1$ to $n$. Hence there will always exist non trivial values of $\alpha_i$ such that Eq. 3.3 is satisfied and the set is linearly dependent.

## 3.2 Basis

Linearly independent vectors have a number of useful properties. An important property concerns using a linearly independent set of vectors to represent other vectors. We will see later that these ideas can be extended to represent functions as well. If a vector $\mathbf{x}$ lies in a finite dimensional space $X$ then we would like to represent $\mathbf{x}$ in a collection of suitable vectors which we will call a basis for the space $X$. A finite collection of vectors $\phi_i$ is said to form a basis for the finite dimensional space $X$ if each vector in $X$ can be represented uniquely as a linear combination of the basis vectors.

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \phi_i = 0 \quad \forall \quad \mathbf{x} \in X \tag{3.4}$$

The term unique in the definition implies that for a given basis set $\phi_i$ and $\mathbf{x}$ the $\alpha_i$ values are uniquely determined Let us illustrate these ideas with some simple examples of basis sets.

**Example 3**: The vectors

$$\phi_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \phi_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \phi_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

form a basis for vectors in $\mathbf{R}^3$, which implies that any vector in $\mathbf{R}^3$ can be represented uniquely using a linear combination of the above vectors. If $a$, $b$ and $c$, represent the components of an arbitrary vector in $\mathbf{R}^3$ then

$$\alpha_1\phi_1 + \alpha_2\phi_2 + \alpha_3\phi_3 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

implies that the coefficients of the expansion are uniquely determined as $\alpha_1 = a$, $\alpha_2 = b$ and $\alpha_3 = c$.

**Example 4**: The vectors given in Example 1 also constitute a basis for vectors in $\mathbf{R}^3$ since the determinant of the resulting matrix formed from the column vectors is non zero.

From the above examples it is clear that for an $n$-dimensional vector space any set of $n$ linearly independent vectors form a suitable basis for the space. In seeking a suitable basis, the representation is complete when the coefficients of the expansion given in Eq. 3.4 are obtained. Clearly some basis sets simplify the determination of these coefficients and the basis in Example 3 was one example of a convenient basis, referred to as the orthonormal basis set. Thus, vectors in a basis are linearly independent and in an $n$ dimensional vector space any set of $n$ linearly independent vectors form a basis for the space

**Dimension of a basis**: The linear space $X$ is $n$ dimensional if it possesses a set of $n$ linearly independent vectors, but every $n + 1$th set is linearly dependent. Equivalently, the number of vectors in a basis is its dimension.

**Example 5:** The set of polynomials of degree $< n$ constitute a basis for an $n$ dimensional linear space of polynomials of degree $< n$. The basis set is

$$\phi_1 = 1\,,\ \phi_2 = x\,,\ \ldots,\ \phi_n = x^{n-1}$$

## 3.3 Linear independence of functions

We next extend the concepts of linear independence for functions. Consider the set of functions, $f_1(x), f_2(x), f_3(x) \ldots f_n(x)$ which are differentiable $n-1$ times on the interval $[a, b]$. The functions are linearly independent on $[a, b]$ if

$$\alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots \alpha_n f_n(x) = 0 \quad \forall\ x \in [a, b] \tag{3.5}$$

implies that $\alpha_i = 0, i = 1 \ldots n$. Differentiating Eq. 3.5 $n-1$ times, a set of equations involving the derivatives of the functions can be generated.

$$\begin{aligned}
\alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \ldots \alpha_n f_n(x) &= 0 \\
\alpha_1 f_1'(x) + \alpha_2 f_2'(x) + \alpha_3 f_3'(x) + \ldots \alpha_n f_n'(x) &= 0 \\
&\ldots\ldots\ldots \\
\alpha_1 f_1^{(n-1)}(x) + \alpha_2 f_2^{(n-1)}(x) + \alpha_3 f_3^{(n-1)}(x) + \ldots \alpha_n f_n^{(n-1)}(x) &= 0
\end{aligned}$$

Eq. 3.6 represents a set of homogeneous equations and the Wronskian is the determinant formed by the functions,

$$|W(f_1(x), f_2(x), f_3(x) \ldots f_n(x)| = \begin{vmatrix} f_1(x) & f_2(x) & f_3(x) & \ldots & f_n(x) \\ f_1'(x) & f_2'(x) & f_3'(x) & \ldots & f_n'(x) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & f_3^{(n-1)}(x) & \ldots & f_n^{(n-1)}(x) \end{vmatrix} \tag{3.6}$$

For Eq. 3.5 to have only the trivial solution, $|W| \neq 0 \quad \forall\, x \in [a, b]$. In this case the set of functions $f_1(x), f_2(x) \ldots f_n(x)$ is said to be linearly independent. However if the Wronskian vanishes for some or all $x \in [a, b]$ it does not necessarily imply that the set is linearly dependent. Thus $|W| \neq 0 \quad \forall\, x \in [a, b]$ is only a *sufficient condition* for the linear independence of the set of functions.

**Example**: $f_1(x) = \sinh x$, $f_2(x) = \cosh x$,

$$|W(f_1(x), f_2(x))| = \begin{vmatrix} \sinh x & \cosh x \\ \cosh x & \sinh x \end{vmatrix} = 1 \neq 0 \tag{3.7}$$

Thus $\sinh x$ and $\cosh x$ constitute a linearly independent set of functions.

**Example**: Consider the polynomials, $f_1(x) = 1$, $f_2(x) = x$ and $f_3(x) = x^2$

$$|W(f_1(x), f_2(x), f_3(x))| = \begin{vmatrix} 1 & x & x^2 \\ 0 & 1 & 2x \\ 0 & 0 & 2 \end{vmatrix} = 2 \neq 0 \tag{3.8}$$

Thus the set of polynomials constitute a linearly independent set of functions. This can be extended a set of $n^{th}$ degree polynomials.

**Example**: $f_1(x) = x^2$, $f_2(x) = 2x^2$,

$$|W(f_1(x), f_2(x))| = \begin{vmatrix} x^2 & 2x^2 \\ 2x & 4x \end{vmatrix} = 0 \tag{3.9}$$

and the set is linearly dependent.

**Example**: $f_1(x) = x$, $f_2(x) = x^2$, $x \in [0, 1]$

$$\alpha_1 x + \alpha_2 x^2 = 0$$

For $x(\alpha_1 + \alpha_2 x) = 0$ for all $x \in [0, 1]$, $\alpha_1 = \alpha_2 = 0$. Thus the set is linearly independent. Upon examining the Wronskian,

$$|W(f_1(x), f_2(x))| = \begin{vmatrix} x & x^2 \\ 1 & 2x \end{vmatrix} = x^2 \tag{3.10}$$

$W$ vanishes for $x = 0$. This is an example where the vanishing of the Wronskian for a particular value of $x$ does not imply that the set is linearly dependent. Clearly this set is linearly independent.

**Example**: $f_1(x) = x^2$, $f_2(x) = x|x|$, $x \in [-1, 1]$

$$\alpha_1 x + \alpha_2 x|x| = 0$$

$\alpha_1 = -\alpha_2|x|/x$. For $-1 < x < 0$, $\alpha_1 = \alpha_2$. For $0 < x < 0$, $\alpha_1 = -\alpha_2$ and at $x = 0$, $\alpha_1, \alpha_2$ are arbitrary. Thus the only way in which $\alpha_1 x + \alpha_2 x|x| = 0$ can be identically zero $\forall\, x$ is when $\alpha_1 = \alpha_2 = 0$. Hence $f_1(x) = x^2$, $f_2(x) = x|x|$, $x \in [-1, 1]$ consitute a linearly independent set. The Wronskian for this case is,

$$|W(f_1(x), f_2(x))| = \begin{vmatrix} x^2 & x|x| \\ 2x & |x| + xh(x) \end{vmatrix} = x^2 \text{ where } h(x) = \frac{d|x|}{dx} = \begin{cases} 1 & 0 < x < 1 \\ -1 & -1 < x < 0 \end{cases}$$

In this case $W = -x^2|x| + x^3 h(x) = 0\, \forall\, x$. This is another example were the vanishing of the Wronskian does not imply that the set is linearly dependent.

## 3.4 Solution of linear equations

One of our primary goals lies in seeking solutions to the general class of linear equations of the form,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{3.11}$$

where $\mathbf{A}$ is in general an $m \times n$ matrix. While discussing issues relating to the solutions of Eq. 3.11 we will make use of the the null space and range space of $\mathbf{A}$ which will also use the ideas of basis sets introduced in this Chapter.

The existence or solvability condition for a set of linear algebraic equations of the form in Eq. 3.11 can be stated as follows. $\mathbf{Ax} = \mathbf{b}$ is solvable if the rank of $\mathbf{A}$ is equal to the rank of the augmented matrix $\mathbf{A}|\mathbf{b}$. The augmented matrix is obtained by adding an extra column vector $\mathbf{b}$ to the matrix $\mathbf{A}$. We illustrate the solvability conditions with reference to the examples of the echelon matrices given in the previous Chapter. Consider the following row reduced forms for the augmented matrices,

$$\left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 2 & 0 \end{array}\right) \quad \left(\begin{array}{cccc|c} 1 & 1 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad \left(\begin{array}{cccc|c} 2 & 1 & 5 & 2 & 1 \\ 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 5 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{3.12}$$

The first and third augmented matrices satisfy the rank criterion and are hence solvable. Once the equations are solvable we inquire into the condition of uniqueness. To answer this we first examine the solutions to the homogeneous problem

$$\mathbf{Ax} = 0 \tag{3.13}$$

and define the null space of $\mathbf{A}$ denoted as $\mathcal{N}(\mathbf{A})$. $\mathcal{N}(\mathbf{A})$ consists of all vectors $\mathbf{x}$ that satisfy the homogeneous equation, Eq. 3.13. We illustrate how the null space can be obtained for the following set of algebraic equations,

$$\begin{aligned} -x_1 + x_3 + 2x_4 &= 0 \\ -x_1 + x_2 - x_4 &= 0 \\ -x_2 + x_3 + 3x_4 &= 0 \\ x_1 - 2x_2 + x_3 + 4x_4 &= 0 \end{aligned} \tag{3.14}$$

Using a series of row operations the matrix can be reduced as follows,

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 1 & 2 \\ -1 & 1 & 0 & -1 \\ 0 & -1 & 1 & 3 \\ 1 & -2 & 1 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.15}$$

resulting in the following two linear equations

$$\begin{aligned} x_1 - x_3 - 2x_4 &= 0 \\ x_2 - x_3 - 3x_4 &= 0 \end{aligned}$$

If $x_3 = \alpha_1$ and $x_4 = \alpha_2$ the solution vector $\mathbf{x}$ can be written in two basis vectors as follows.

$$\mathbf{x} = \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 3 \\ 0 \\ 1 \end{pmatrix}$$

We make a few observations. As a check on the solution procedure we should ensure that $\mathbf{x}$ given above satisfies the original set of equations. The rank of the matrix in the above example is 2 which is equal to the number of linearly independent equations. Since the number of unknowns is 4 the dimension of $\mathcal{N}(\mathbf{A})$ is 4 - 2 = 2. The dimension of $\mathcal{N}(\mathbf{A})$ is simply the number of linearly independent vectors in the basis used to represent the solution space $\mathbf{x}$. This result is easily generalizable. For a general $m \times n$ matrix whose rank is $r$ the dimension of $\mathcal{N}(\mathbf{A})$ is $n - r$. Note that $n$ is the number of unknowns in the set of equations and $n - r$ which is the number of arbitrary ways in which the unknowns can be specified yields the dimension of the basis. Clearly there is no unique way of choosing the unknowns and hence the basis for $\mathcal{N}(\mathbf{A})$ is not unique. However the dimension of $\mathcal{N}(\mathbf{A})$ is fixed. $\mathcal{N}(\mathbf{A})$ is empty when, $n = r$ then the only solution to the homogeneous problem is the trivial solution. This leads to an important result.

**Theorem**: If $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution, then $\mathbf{Ax} = \mathbf{b}$ has a unique solution.

**Proof**: Let the inhomogeneous equation $\mathbf{Ax} = \mathbf{b}$, have two solutions $\mathbf{u}$ and $\mathbf{v}$. Then

$$\begin{aligned} \mathbf{Au} &= \mathbf{b} \\ \mathbf{Av} &= \mathbf{b} \end{aligned}$$

Subtracting the two equations

$$\mathbf{Aw} = 0 \tag{3.16}$$

where $\mathbf{w} = \mathbf{u} - \mathbf{v}$. Since $\mathbf{Ax} = 0$ has only the trivial solution $\mathbf{w} = 0$ and $\mathbf{u} = \mathbf{v}$. Hence $\mathbf{Ax} = \mathbf{b}$ has a unique solution. The above proof is always true if the matrix is square. The proof is true for any $m \times n$ matrix, provided the inhomogeneous equation $\mathbf{Ax} = \mathbf{b}$ is solvable. Example 6 in this section illustrates this situation. Later we will see that a similar proof can be used for some linear differential and integral operators. If the matrix is square and the inverse exists (determinant of $\mathbf{A} \neq 0$ or equivalently rank, $r = n$), then $\mathbf{Ax} = \mathbf{b}$ has a unique solution which is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Further the solution exists for any vector $\mathbf{b}$. The last statement is equivalent to

noting that for a general nonsingular matrix $n \times n$ whose rank = $n$, the rank of the augmented matrix must also equal $n$ and is consistent with our solvability conditions based on the notion of the rank.

Earlier we saw that a basis could be defined for the null space of $\mathbf{A}$. Using the solvability condition based on rank equivalence, we can define an additional space relevant to understanding the solutions to linear equations as the range space of $\mathbf{A}$ also denoted as $\mathcal{R}(\mathbf{A})$. $\mathcal{R}(\mathbf{A})$ consists of all vectors such that $\mathbf{Ax} = \mathbf{b}$ is solvable. We illustrate this with a simple example. Consider the augmented matrix where $b_1$ and $b_2$ represent elements of vector $\mathbf{b}$.

$$\begin{pmatrix} 2 & 3 & | & b_1 \\ 6 & 9 & | & b_2 \end{pmatrix} \quad \rightarrow \begin{pmatrix} 2 & 3 & | & b_1 \\ 0 & 0 & | & 3b_1 - b_2 \end{pmatrix} \tag{3.17}$$

The second matrix is obtained by elementary row operations. The solvability condition requires that $3b_1 - b_2 = 0$ resulting in the following basis for $\mathcal{R}(\mathbf{A})$,

$$\mathbf{b} = \alpha \begin{pmatrix} 1 \\ 3 \end{pmatrix} \tag{3.18}$$

where $\alpha$ is an arbitrary scalar. In the above example, dim[$\mathcal{R}(\mathbf{A})$] = 1 $\equiv$ r. Using the definition of $\mathcal{R}(\mathbf{A})$, the solvability condition is equivalent to stating that $\mathbf{Ax} = \mathbf{b}$ is solvable if $\mathbf{b}$ lies in $\mathcal{R}(\mathbf{A})$.

To complete the solution scenario for the linear equations we need to discuss the situation when the homogeneous equation, $\mathbf{Ax} = 0$ has non-trivial solutions i.e when $\mathcal{N}(\mathbf{A})$ is not empty.

**Theorem**: If $\mathbf{Ax} = 0$ has non-trivial solutions then $\mathbf{Ax} = \mathbf{b}$ may or may not be solvable. If it is solvable then it has an infinity of solutions.

If $\mathbf{A}$ has non-trivial solutions then the rank, $r < n$ for a $n \times n$ square matrix and for an $(m \times n)$ matrix, $r < n$ for both $m < n$ and $m > n$. If the homogeneous problem has non-trivial solutions then $\mathbf{Ax} = \mathbf{b}$ is solvable if and only if the rank of the matrix equals the rank of the augmented matrix. If this solvability condition is satisfied, then a solution exists, and the system has an infinity of solutions. The infinity of solutions is due to the non trivial solutions of the homogeneous problem and hence can be related to $\mathcal{N}(\mathbf{A})$. The solution to $\mathbf{Ax} = \mathbf{b}$ can in general be split into two parts in the following manner,

$$\mathbf{x} = \sum_{i=1}^{k} \alpha_i \phi_i + \mathbf{x}_p. \tag{3.19}$$

The first term on the right hand side represents part of the solution that lies in $\mathcal{N}(\mathbf{A})$ whose dimension (without loss of generality) is assumed to be $k$, and $\phi_i$ form the basis for $\mathcal{N}(\mathbf{A})$. $\mathbf{x}_p$ is a particular solution to $\mathbf{Ax} = \mathbf{b}$. To show that $\mathbf{x}$ given in Eq. 3.19 is a general solution, we operate on $\mathbf{x}$ with $\mathbf{A}$. Hence

$$\begin{aligned}\mathbf{Ax} &= \mathbf{A}(\sum_{i=1}^{k} \alpha_i \phi_i) + A\mathbf{x}_p \\ &= (\sum_{i=1}^{k} \alpha_i \mathbf{A}\phi_i) + A\mathbf{x}_p \\ &= \mathbf{b}\end{aligned}$$

We note that since $\phi_i$ forms the basis for $\mathcal{N}(\mathbf{A})$, $\mathbf{A}\phi_i = 0$. The infinity of solutions is due to solutions in $\mathcal{N}(\mathbf{A})$ since $\alpha_i$ are arbitrary scalars. If the only solution to $\mathbf{Ax} = 0$ is the trivial solution then $\mathcal{N}(\mathbf{A})$ is empty and the solution is unique. In this case $\mathbf{x} = \mathbf{x}_p$ assuming that the solvability condition is satisfied. The existence and uniqueness conditions for $\mathbf{Ax} = \mathbf{b}$ discussed above are summarized in Figure 3.1.

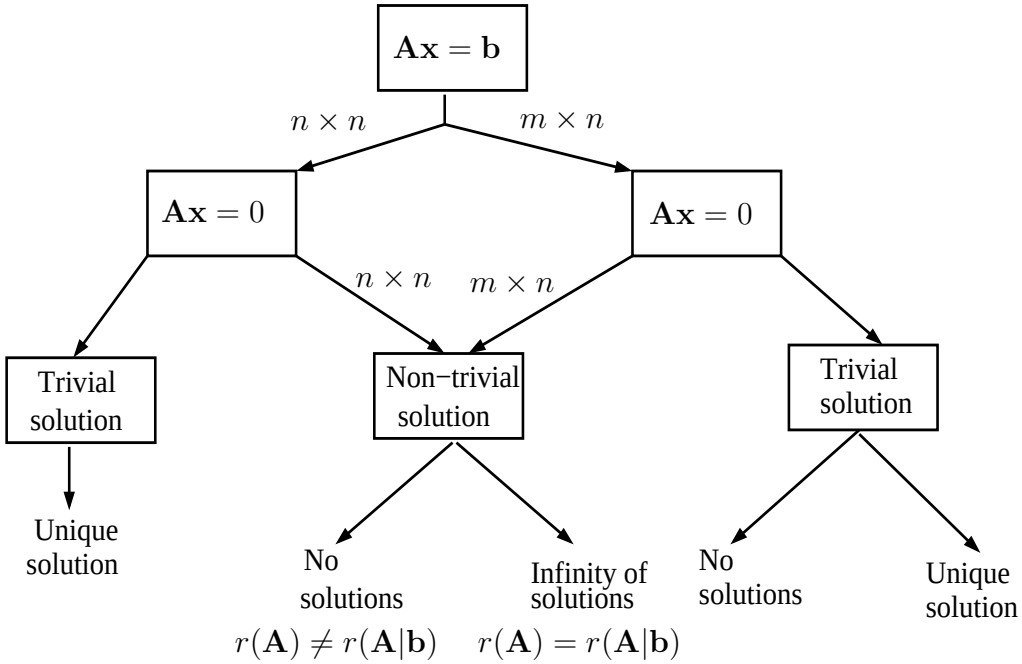



Figure 3.1: Illustration of the various solution scenarios that are encountered while solving linear equations.

We end this section on solutions to linear equations with a geometric interpretation of

the different solution scenarios discussed above.

### 3.4.1 Geometrical Interpretation

Consider the following set of linear algebraic equations with two unkowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned} \tag{3.20}$$

Solutions to the above equations can be analyzed by plotting $x_2$ vs $x_1$ on a two dimensional plot as shown in the Figures below. We assume that $a_{11}/a_{12} > 0$ and $a_{22}/a_{21} > 0$. Hence both lines in the above equations will have negative slopes.

**Case 1**: The determinant, $a_{11}a_{22} - a_{21}a_{12}$ is non-zero. Hence $\mathbf{Ax} = \mathbf{0}$ has the trivial solution. This is illustrated in Fig. 3.2, where the solutions to $\mathbf{Ax} = \mathbf{0}$ is only the trivial solution indicated by the intersection of the two lines at the origin. In this situation both lines have different slopes. Further, Eq. 3.20 has a unique solution for any vector $\mathbf{b}$ lying in the plane.

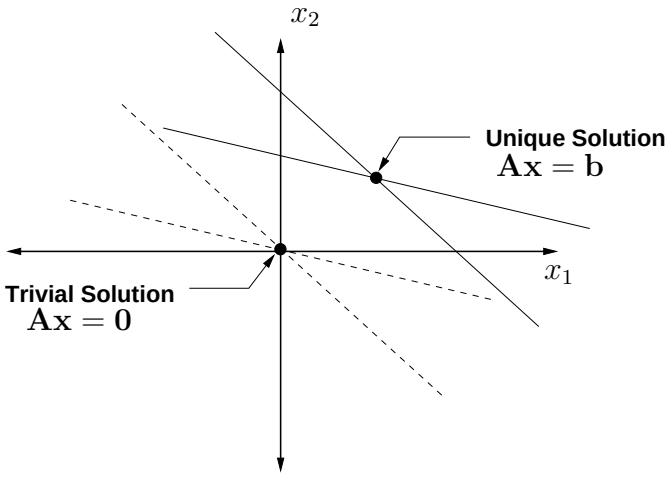



Figure 3.2: Solution of linear equations illustrating a unique solution. The dashed lines represent solutions to $\mathbf{Ax} = \mathbf{0}$.

**Case 2:** The determinant, $a_{11}a_{22} - a_{21}a_{12}$ is zero. This implies that both lines have the same slopes (Fig. 3.3). Hence $\mathbf{Ax} = \mathbf{0}$ has an infinity of solutions indicated by the dashed line that

passes through the origin. If the solvability criterion (rank condition) is satisfied then the solution to $\mathbf{Ax} = \mathbf{b}$ consists of all points on a line having the same slope with intercept $b_1/a_{12}$

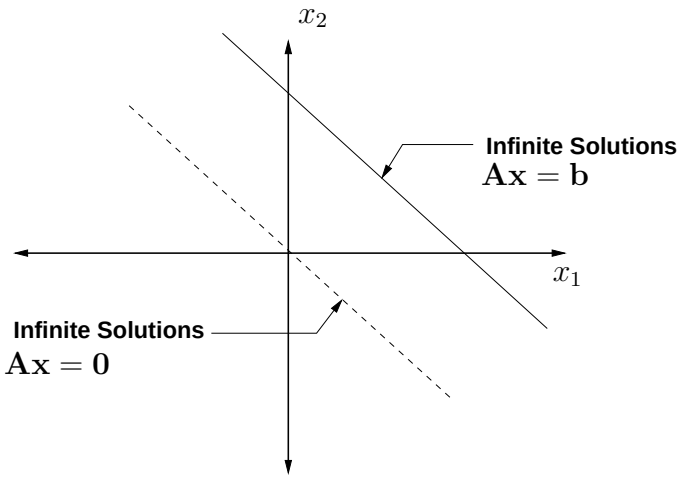



Figure 3.3: Solution of linear equations illustrating an infinity of solutions. The dashed line represents the solutions to $\mathbf{Ax} = \mathbf{0}$.

**Case 3:** The determinant, $a_{11}a_{22} - a_{21}a_{12}$ is zero. This implies that both lines have the same slopes. Hence as in Case 2, $\mathbf{Ax} = \mathbf{0}$ has an infinity of solutions indicated by the dashed line that passes through the origin. If the solvability criterion (rank condition) is not satisfied then $\mathbf{Ax} = \mathbf{b}$ does not have a solution as illustrated in Fig. 3.4.

**Example 6**:

Let us examine the solvability conditions for the set of linear equations,

$$\begin{aligned} x_1 + 2x_2 &= b_1 \\ 2x_1 + 4x_2 &= b_2 \\ x_1 &= b_3 \end{aligned} \tag{3.21}$$

It is easy to see that the only solution to the homogeneous equation $\mathbf{Ax} = \mathbf{0}$ is the trivial solution. Hence $\mathcal{N}(\mathbf{A})$ is empty. The range space $\mathbf{A}$ consists of,

$$\mathbf{b} = \alpha_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

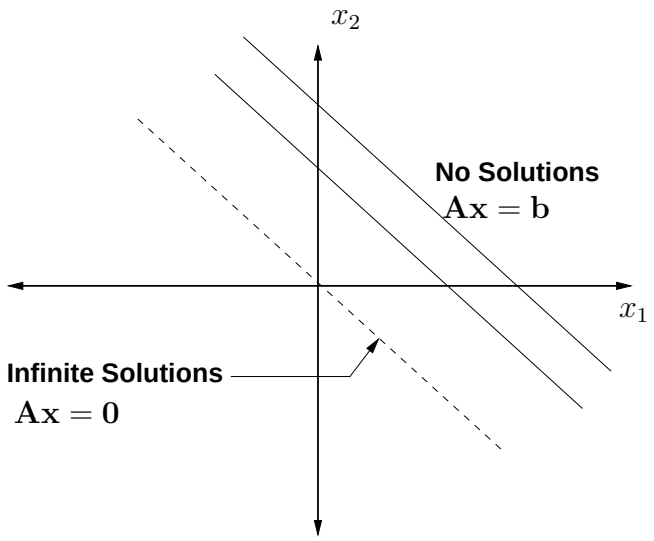



Figure 3.4: Solution of linear equations illustrating no solutions. The dashed line represents the solutions to $\mathbf{Ax} = \mathbf{0}$.

Hence $2b_1 = b_2$ and $b_3$ is arbitrary. The solutions are,

$$x_1 = b_3, \qquad x_2 = \frac{b_1 - b_3}{2} \equiv \frac{b_2 - 2b_3}{4}$$

This is an illustrative example, as it is a situation of an $m \times n$ system where the null space is empty. If the $\mathbf{b}$ lies in the range, then the system of equations has a unique solution. Figure 3.5 graphically illustrates some possible solution scenarios.

## 3.5 Summary

Starting from the definitions of the linear or vector space, we introduced the concept of linear independence and subsequently notions of basis sets and dimensions of basis. The idea of representing vectors or functions in a suitable basis has far reaching consequences in functional analysis and solutions of differential equations. In this Chapter we saw how a basis could be used to construct the null space and range space of a matrix and connect the dimensions of these spaces to the now familiar definition of the rank of the matrix. The theorems on solutions of linear systems completes the discussion on existence and uniqueness for this class of inhomogeneous equations which can be represented as $\mathbf{Ax} = \mathbf{b}$. The starting point for the analysis was to investigate the solutions of the homogeneous system. Figure 3.1 schematically illustrates the

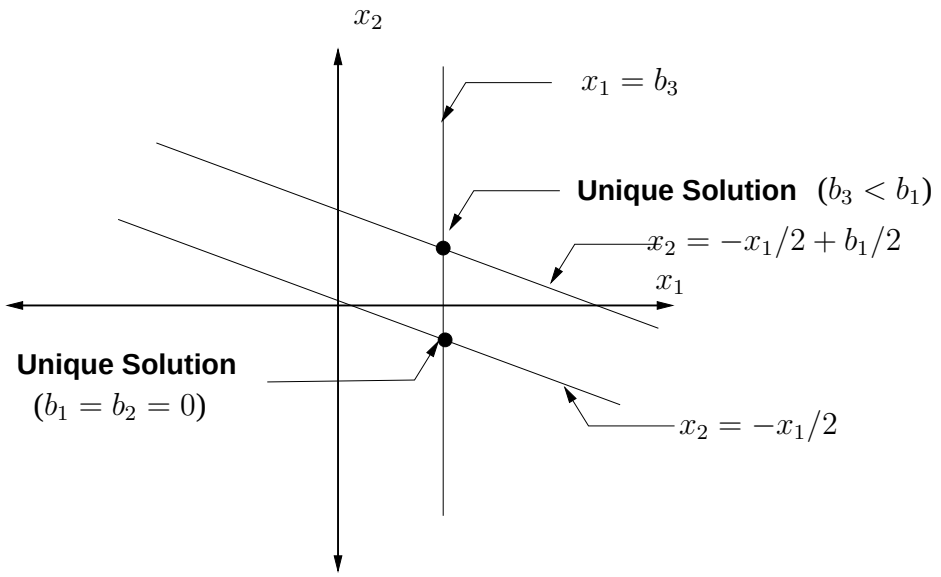



Figure 3.5: Solution of linear equations illustrating two possible solution scenarios for the set of linear equations given in Example 6. In one case, $b_3 < b_1$ and in the second case $b_1 = b_2 = 0$. In both cases the solutions are points obtained with the intersection by the vertical line $x_1 = b_3$.

various scenarios

Since we are interested in existence and uniqueness conditions for ordinary differential equations, we have to abandon the notions of ranks and determinants that form the basic tools to analyse a linear system of equations. We begin to develop a more complete theory of linear operators in the next Chapter by introducing the inner product space and the adjoint operator. Once we are equipped with this formalism to study linear differential equations later in the book, we will first revisit the theorems developed in this Chapter to understand the generality and utility of these tools and ideas.

## PROBLEMS

1. Which of the following column vectors can be used to construct a basis for the three dimensional vector space $\mathbf{R^3}$.

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$$

   Once you have picked an appropriate basis set, show that any vector in $\mathbf{R^3}$ can be uniquely represented using this basis. In other words show that the vectors you have chosen form a valid basis for $\mathbf{R^3}$.

2. Consider the space $X$ consisting of all polynomials, $f(x), a \leq x \leq b$, with real coefficients and degree not exceeding $n$.

   (a) Show that $X$ is a real linear (vector) space.

   (b) What is the dimension of this space ?

   (c) Define a suitable basis for this space of polynomials.

   (d) Show that your basis does constitute a linearly independent set of vectors.

3. Consider the following functions

$$\phi_n = (1-t)^{n-1}$$

   for $n = 1$ to $4$.

   (a) Do these form a linearly independent set ?

   (b) What is the dimension of the vector space they span ?

   (c) Using these functions construct a basis to represent the polynomial $3t^3 - 2t^2 + 6t - 5$. Find the coefficients of the expansion.

4. Show that the presence of a zero vector in a set of linearly independent vectors makes the set of vectors linearly dependent.

5. Consider the following matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 1 & 6 \\ 1 & 2 & 4 \end{pmatrix}$$

(a) Reduce $\mathbf{A}$ to its echelon form.

(b) What is the rank of $\mathbf{A}$?

(c) What is the dimension of the null space, $\mathcal{N}(\mathbf{A})$ of $\mathbf{A}$? Find a basis for $\mathcal{N}(\mathbf{A})$.

(d) What is the dimension of the range space, $\mathcal{R}(\mathbf{A})$ of $\mathbf{A}$? Find a basis for $\mathcal{R}(\mathbf{A})$.

(e) Using your answer from part (f) identify which of the following vectors $\mathbf{b}$ will yield a solution to $\mathbf{Ax} = \mathbf{b}$

$$\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix} \quad \begin{pmatrix} 6 \\ 5 \\ 12 \end{pmatrix}$$

(f) Find the solutions to $\mathbf{Ax} = \mathbf{b}$ for those vectors $\mathbf{b}$ in part (g) for which solutions are feasible. Note that your solutions consist of a vector that belongs to the null space of $\mathbf{A}$ and a vector that satisfies $\mathbf{Ax} = \mathbf{b}$.

6. Consider the following set of linear algebraic equations

$$\begin{aligned} x_1 + 2x_2 + x_3 + 2x_4 - 3x_5 &= 2 \\ 3x_1 + 6x_2 + 4x_3 - x_4 - 2x_5 &= -1 \\ 4x_1 + 8x_2 + 5x_3 + x_4 - x_5 &= 1 \\ -2x_1 - 4x_2 - 3x_3 + 3x_4 - 5x_5 &= 3 \end{aligned}$$

(a) Reduce to echelon form.

(b) Find the basis for null space of $\mathbf{A}$.

(c) Find the basis for the range of $\mathbf{A}$.

(d) Construct the complete solution to the set of equations.

(e) Does the system have a unique solution? If not, how many solutions does the system possess?

7. Consider the following matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 1 & 2 \\ 4 & -1 & 9 \\ 2 & 1 & 1 \end{pmatrix}$$

   (a) What is the rank of $\mathbf{A}$?

   (b) What is the dimension of the null space $(\mathcal{N}(\mathbf{A}))$ of $\mathbf{A}$?

   (c) What is the dimension of the range space $(\mathcal{R}(\mathbf{A}))$ of $\mathbf{A}$? Find a basis for $\mathcal{R}(\mathbf{A})$.

   (d) Next consider the transpose of the matrix $\mathbf{A}$. Find a basis for the null space of $\mathbf{A}^T$.

   (e) Construct a vector space such that the vectors in the space are orthogonal to the null space of $\mathbf{A}^T$. What is the dimension of this vector space ? Compare this orthogonal vector space with $\mathcal{R}(\mathbf{A})$. Can you draw any conclusions.

   (f) Find the solutions to $\mathbf{Ax} = \mathbf{b}$ for

$$\mathbf{b} = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 2 \end{pmatrix}$$

   Does the system have a unique solution and why? Illustrate your solution graphically. *Note: This problem is connected with the general Fredholms Alternative theorems to be introduced in the Chapter 4, Sec 4.5*

8. Determine the ranks, dimensions and suitable basis for both the $(\mathcal{N}(\mathbf{A}))$ and $(\mathcal{R}(\mathbf{A}))$ for the following sets of linear algebraic equations. If the right hand side vector $\mathbf{b}$ is given, obtain a particular solution to the set of equations.

   (a)

$$\begin{aligned} x_1 - x_2 + 3x_3 + 2x_4 &= b_1 \\ 3x_1 + x_2 - x_3 + x_4 &= b_2 \\ -x_1 - 3x_2 + 7x_3 + 3x_4 &= b_3 \end{aligned}$$

(b)

$$\begin{aligned} x_1 + 2x_2 - x_3 + x_5 &= b_1 \\ 3x_1 + 2x_2 + x_4 &= b_2 \\ x_1 - 2x_2 + 2x_3 + x_4 - 2x_5 &= b_3 \end{aligned}$$

(c)

$$\begin{aligned} 5x_1 + 10x_2 + x_3 - 2x_4 &= 6 \\ -x_1 + x_2 - 2x_3 + x_4 &= 0 \\ 2x_1 + 3x_2 + x_3 - x_4 &= 2 \\ 6x_1 + 9x_2 + 3x_3 - 3x_4 &= 6 \end{aligned}$$

(d)

$$\begin{aligned} x_1 + x_2 - x_3 &= b_1 \\ -2x_1 - x_2 + x_3 &= b_2 \\ x + 2x_2 - 2x_3 &= b_3 \end{aligned}$$

# Chapter 4

# Inner Products, Orthogonality and the Adjoint Operator

While defining the linear space or vector space we were primarily concerned with elements or vectors that conform to the rules of addition and scalar multiplication. These are algebraic properties. The simplicity of the linear space was sufficient to introduce ideas such as linear independence and basis sets in a finite dimensional setting. We observe that notions of distance, length and angles between the elements of the space, which reflect geometric properties were not discussed.

In this chapter we define the inner product space which provides the necessary framework to introduce geometric properties. The primary motivation for this is to lay the grounds for discussing orthogonality and its relationship to representation of vectors or functions in a suitable basis set. In this Chapter the Gram-Schmidt orthogonalization process and its relationship to well known orthogonal polynomials, such as the Legendre and Hermite polynomials will be developed. The inner product space allows us to introduce the Schwarz and triangular inequalities. We end this chapter with the definition of the adjoint operator and its utility in studying issues of uniqueness and existence of non-homogeneous linear equations,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4.1}$$

## 4.1 Inner Product Spaces

The inner product space consists of a linear space $X$ on which the inner product, denoted by $< \cdot, \cdot >$ is defined, where the dots represent any two elements in the space. If $u$, $v$ and $w$ are contained in $X$, and $\alpha$ is an arbitrary scalar contained with the scalar field associated with $X$, then the inner product satisfies the following axioms

1. Linearity:

$$< u + v, w >=< u, w > + < v, w >$$

and

$$< \alpha u, v >= \alpha < u, v >$$

2. Symmetry

$$< u, v >= \overline{< v, u >}$$

3. Positive Definiteness

$$< u, u > \quad > \quad 0 \qquad \text{when} \qquad u \neq 0$$

4.

$$< u, \alpha v >= \overline{\alpha} < u, v >$$

In the above definitions the overbar is used to denote the complex conjugate. Note that the inner product always results in a scalar quantity.

The inner product inherently contains the definition of the length or norm, denoted by $\| \cdot \|$. The norm of $u$ is related to its inner product in the following manner

$$\|u\|^2 =< u, u > \tag{4.2}$$

**Example 1**

Consider two vectors **u** and **v** in an $n$-dimensional vector space. The inner product,

$$< \mathbf{u}, \mathbf{v} > \quad = \quad u_1\overline{v_1} + u_2\overline{v_2} + \ldots u_n\overline{v_n} \tag{4.3}$$

$$= \quad \sum_i^n u_i\overline{v_i} \tag{4.4}$$

The norm of the vector $\mathbf{u}$,

$$\|\mathbf{u}\| = \sqrt{< u, u >} = \sqrt{\sum_{i=1}^{n} u_i \overline{u_i}}$$

The use of the complex conjugate while defining the inner product is consistent with our notion of the length of a vector in the complex plane. Consider the point with co-ordinates $(1, i)$ in the complex plane denoted by the vector

$$\mathbf{u} = \begin{pmatrix} 1 \\ i \end{pmatrix} \tag{4.5}$$

If one were to use the definition of the inner product in the absence of the complex conjugate then it would imply that a non-zero vector has a zero length! Using the definition of the inner product given in Eq. 4.4, the norm of the vector given in Eq. 4.5, $\|\mathbf{u}\| = \sqrt{2}$.

*Question: Show that the inner product as given in Eq. 4.4 satisfies the axioms of the inner product space. Thus the vectors of the $n$-dimensional vector space form an inner product space.*

**Example 2**

Consider two functions $f(x)$ and $g(x)$ which belong to the space of continuous functions with $x \in [a, b]$. The inner product between the two functions,

$$< f(x), g(x) >= \int_a^b f(x)\overline{g(x)}\, dx$$

The square of the norm,

$$\begin{aligned} \|f(x)\|^2 &= \int_a^b f(x)\overline{f(x)}\, dx \\ &= \int_a^b |f(x)|^2\, dx \end{aligned} \tag{4.6}$$

If $f(x) = x$ and $g(x) = e^{-ix}$ then

$$< f(x), g(x) >= \int_a^b x e^{ix}\, dx$$

Further

$$< g(x), f(x) >= \int_a^b e^{-ix} x\, dx$$

and it is easily seen that $< f(x), g(x) >$ = $\overline{< g(x), f(x) >}$, thereby satisfying the symmetry property.

In the above examples it is easy to show that the definitions of the inner product satisfy the axioms of the inner product space. One might naturally inquire, if there are alternate definitions of the inner product. Indeed, other definitions do exist and we will encounter some of them later in the book. However, as long as the definition of the inner product satisfies the axioms of the inner product space it is a valid candidate.

## 4.2 Orthogonality

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be orthogonal if their inner product is identically zero,

$$< \mathbf{u}, \mathbf{v} >= \sum_{i}^{n} u_i \overline{v_i} = 0$$

Similarly two functions $f(x)$ and $g(x)$ are orthogonal on the interval $x \in [a, b]$ if

$$< f(x), g(x) >= \int_a^b f(x)\overline{g(x)}, dx = 0$$

The collection of vectors $\mathbf{u}_1, \mathbf{u}_2 \ldots, \mathbf{u}_n$ are said to form an orthogonal set if

$$< \mathbf{u}_i, \mathbf{u}_j >= 0 \quad \text{if} \quad i \neq j$$

and the set is said to be orthonormal if

$$< \mathbf{u_i}, \mathbf{u_j} >= \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{4.7}$$

The norm of each vector in an orthonormal set is unity. Hence an orthonormal set is obtained from an orthogonal set by dividing each vector by its length or norm.

**Example 3**

Consider the vectors

$$\mathbf{u_1} = \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix}, \mathbf{u_2} = \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix}, \mathbf{u_3} = \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}$$

These form an orthogonal set, since $< \mathbf{u_i}, \mathbf{u_j} >= 0$ when $i \neq j$. The corresponding orthonormal set obtained by dividing each of the above vectors, $\mathbf{u_i}$ by its norm, $\|\mathbf{u_i}\|$, is the familiar set of unit vectors which constitute a basis in $\mathcal{R}^3$

$$\mathbf{e_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e_2} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e_3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

**Example 4**

The set of functions, $u_1(x) = \sin \pi x, u_2(x) = \sin 2\pi x, \ldots u_n(x) = \sin n\pi x$ form an orthogonal set in the interval $0 \leq x \leq 1$. Hence

$$< u_n(x), u_m(x) >= \int_0^1 \sin n\pi x \, \sin m\pi x \, dx = \begin{cases} 0 & m \neq n \\ 1/2 & m = n \end{cases}$$

The corresponding orthonormal set, $\left\{v_n(x) = \sqrt{2} \sin n\pi x\right\}$.

## 4.3 Orthogonality and Basis Sets

Perhaps the most elegant and useful property of an orthonormal set is the utility as a basis to represent other vectors or functions. Consider representing a vector $\mathbf{x}$ in a finite dimensional space using a suitable basis, $\{\phi_i\}$. We will first assume that the basis does not form an orthogonal set and is simply a linearly independent set. Let $\mathbf{x}$ be a vector in the complex plane, $\mathcal{C}^n$.

$$\mathbf{x} = \sum_{i=1}^{N} \alpha_i \phi_i \tag{4.8}$$

To find the coefficients $\alpha_i$ in the expansion, take the inner product of Eq. 4.8 with $\phi_j$. This yields

$$< \mathbf{x}, \phi_j >= \sum_{i=1}^{N} < \alpha_i \phi_i, \phi_j > \qquad j = 1 \ldots N \tag{4.9}$$

The procedure of taking inner products generates a set of $N$ linear algebraic equations which can compactly be written in matrix vector notation as,

$$\mathbf{A}\alpha = \mathbf{b}$$

where $\alpha$ is the vector of unknown coefficients in the expansion, Eq. 4.8 and

$$\begin{aligned} a_{ij} &= < \phi_j, \phi_i > \equiv \overline{< \phi_i, \phi_j >} \\ b_i &= < \mathbf{x}, \phi_i > \end{aligned}$$

If the basis forms an orthonormal set as defined in Eq. 4.7 then the solution is greatly simplified. The matrix $\mathbf{A}$ reduces to the Identity matrix and the solution, which are the coefficients in the expansion (Eq. 4.8)

$$\alpha_i =< \mathbf{x}, \phi_i > \qquad i = 1 \ldots N$$

The above procedure of obtaining the coefficients is similar in function space as well, with the appropriate definition of the inner product. We observe that if we had a basis that was not orthogonal then the procedure results in obtaining a solution to a set of linear algebraic equations. In the case of functions the elements of the resulting coefficient matrix, $\mathbf{A}$ consist of integrals that have to be evaluated. We illustrate the above procedure with examples in both vector and function spaces.

**Example 5**

Consider vectors in $\mathcal{R}^2$.

$$\mathbf{x} = \alpha_1\phi_1 + \alpha_2\phi_2 \tag{4.10}$$

where

$$\mathbf{x} = \begin{pmatrix}1\\2\end{pmatrix}, \qquad \phi_1 = \begin{pmatrix}2\\1\end{pmatrix}, \qquad \phi_2 = \begin{pmatrix}1\\1\end{pmatrix}$$

The resulting set of linear equations can be solved using standard methods. However, in what follows we utilize inner products as illustrated in Sec. 4.3 to obtain coefficients in the expansion given in Eq. 4.10. Taking the inner product of the expansion, with $\phi_1$ and $\phi_2$ respectively, the coefficients in the expansion are obtained by solving the following linear equations,

$$\begin{aligned} 5\alpha_1 + 3\alpha_2 &= 4 \\ 3\alpha_1 + 2\alpha_2 &= 3 \end{aligned}$$

whose solution yields $\alpha_1 = -1$ and $\alpha_2 = 3$. If we use the following orthogonal basis

$$\phi_1 = \begin{pmatrix}-1\\1\end{pmatrix}, \qquad \phi_2 = \begin{pmatrix}1\\1\end{pmatrix}$$

then

$$\begin{aligned} \alpha_1 &= \frac{< \mathbf{x}, \phi_1 >}{< \phi_1, \phi_1 >} = 1/2 \\ \alpha_2 &= \frac{< \mathbf{x}, \phi_2 >}{< \phi_2, \phi_2 >} = 3/2 \end{aligned}$$

Finally, if we use the corresponding orthonormal basis, obtained by normalizing the orthogonal set above,

$$\phi_{\mathbf{1}} = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix}, \qquad \phi_{\mathbf{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

and the coefficients are

$$\begin{aligned} \alpha_1 &= < \mathbf{x}, \phi_1 >= 1/\sqrt{2} \\ \alpha_2 &= < \mathbf{x}, \phi_2 >= 3/\sqrt{2} \end{aligned}$$

The above example illustrates the simplification in the analysis obtained by using an orthonormal basis set, over a basis that is simply linearly independent or even orthogonal.

**Example 6**

Expansion in basis sets is of central importance in functional approximation using Fourier series. Consider representing a function, $f(x)$ for $0 \leq x \leq$ in an infinite $\sin$ series which was shown to form an orthogonal set in Example 4 above.

$$f(x) = \sum_{n=1}^{\infty} a_n \sin n\pi x$$

Taking inner products with $\sin m\pi x$,

$$< f(x), \sin m\pi x >= \sum_{n=1}^{\infty} a_n < \sin n\pi x, \sin m\pi x >$$

Replacing the inner product with integrals,

$$\int_0^1 f(x) \sin m\pi x \, dx = \sum_{n=1}^{\infty} a_n \int_0^1 \sin n\pi x, \sin m\pi x \, dx \qquad m = 1 \ldots n$$

Using the orthogonality property of the functions, $\sin n\pi x,\ n = 1 \ldots \infty$ given in Example 4 above, the expression for the coefficient reduces to,

$$a_n = 2 \int_0^1 f(x) \sin n\pi x \, dx$$

The above expression is obtained by noting that for every $m$ in the previous equation, only the $m^{th}$ term in the expansion survives. We will encounter similar expansions while solving PDEs with the separation of variables technique. The representation of functions in a series expansion of orthonormal sets forms the key foundation for solving PDEs and the central ideas of functional representation presented in this Chapter should be mastered at this point.

## 4.4 Gram-Schmidt Orthogonalization

The Gram-Schmidt (GS) orthogonalization provides a systematic method of constructing an orthogonal set from a linearly independent set of vectors. Given a set of $n$ linearly independent vectors $\{\mathbf{u}_i\}$ the GS process can be used to construct, $\{\mathbf{v}_i\}$, the orthogonal set. Let $\{\mathbf{x}_i\}$ denote the corresponding orthonormal set.

$$\mathbf{v}_1 = \mathbf{u}_1 \qquad \text{and} \qquad \mathbf{x}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$$

Construct the next vector $\mathbf{v}_2$ as a linear combination of $\mathbf{u}_2$ and $\mathbf{x}_1$,

$$\mathbf{v}_2 = \mathbf{u}_2 - \alpha_1 \mathbf{x}_1$$

such that the orthogonality condition $< \mathbf{v}_2, \mathbf{x}_1 >= 0$ is satisfied. Taking inner products of the above equation with $\mathbf{x}_1$, $\alpha_1 =< \mathbf{u}_2, \mathbf{x}_1 >$. Hence

$$\mathbf{v}_2 = \mathbf{u}_2 - < \mathbf{u}_2, \mathbf{x}_1 > \mathbf{x}_1 \quad \text{and} \quad \mathbf{x}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$$

Proceeding in a similar manner

$$\mathbf{v}_3 = \mathbf{u}_3 - \alpha_2 \mathbf{x}_2 - \alpha_3 \mathbf{x}_1$$

Setting $\mathbf{v}_3$ orthogonal to the $\mathbf{x}_2$ and $\mathbf{x}_1$, i.e. $< \mathbf{v}_3, \mathbf{x}_2 >= 0$ and $< \mathbf{v}_3, \mathbf{x}_1 >= 0$ results in $\alpha_2 =< \mathbf{u}_3, \mathbf{x}_2 >$ and $\alpha_3 =< \mathbf{u}_3, \mathbf{x}_1 >$. Hence

$$\mathbf{v}_3 = \mathbf{u}_3 - < \mathbf{u}_3, \mathbf{x}_2 > \mathbf{x}_2 - < \mathbf{u}_3, \mathbf{x}_1 > \mathbf{x}_1 \quad \text{and} \quad \mathbf{x}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|}$$

Continuing in this manner,

$$\mathbf{v}_n = \mathbf{u}_n - < \mathbf{u}_n, \mathbf{x}_{n-1} > \mathbf{x}_{n-1} - < \mathbf{u}_n, \mathbf{x}_{n-2} > \mathbf{x}_{n-2} \ldots, - < \mathbf{u}_n, \mathbf{x}_1 > \mathbf{x}_1$$

and

$$\mathbf{x}_n = \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|}$$

It is easy to show that $< \mathbf{v}_n, \mathbf{v}_m >= 0$ for $m < n$. Hence the set $\{\mathbf{v}_i\}$, is an orthogonal set and $\{\mathbf{x}_i\}$ is an orthonormal set. We note that the above procedure does not depend on the initial ordering of the set of linearly independent vectors, $\{\mathbf{u}_i\}$. However, each ordering will

result in a different set of orthonormal vectors. Hence for a given vector space there are a large number of orthonormal sets. It is easy to visualize this in two dimensions, where two orthogonal vectors in the plane can be rotated by by an arbitrary angle to generate an infinite combination of orthogonal vectors. We illustrate the GS procedure with some examples.

**Example 7**: Consider the set of linearly independent vectors in $\mathcal{R}^2$.

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}$$

Using the GS procedure outlined above,

$$\mathbf{x}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

We next construct $\mathbf{v}_2$

$$\mathbf{v}_2 = \mathbf{u}_2 - < \mathbf{u}_2, \mathbf{x}_1 > \mathbf{x}_1 = \begin{pmatrix} i \\ 1 \end{pmatrix} - i \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv \mathbf{x}_2$$

However if we reorder the initial set of two vectors, such that

$$\mathbf{u}_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

then the resulting orthonormal set is

$$\mathbf{x}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}, \mathbf{x}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$

This example illustrates the non-uniqueness in the orthogonal set obtained using the GS procedure.

**The Schwarz Inequality**

Consider the definition of the dot product of two vectors,

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

where $\theta$ is the angle between the two vectors and is defined in terms of the dot product and the norms of the two vectors. Using the inner product notation

$$\begin{aligned} < \mathbf{u}, \mathbf{v} > &= \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta \\ | < \mathbf{u}, \mathbf{v} > | &= \|\mathbf{u}\| \|\mathbf{v}\| |\cos\theta| \end{aligned}$$

Since $0 \leq |\cos\theta| \leq 1$

$$| < \mathbf{u}, \mathbf{v} > | \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

This is known as the Schwarz inequality. We present a more general derivation below

$$\begin{aligned} 0 &\leq < \mathbf{u} + \alpha\mathbf{v}, \mathbf{u} + \alpha\mathbf{v} > \\ &= < \mathbf{u}, \mathbf{u} + \alpha\mathbf{v} > + < \alpha\mathbf{v}, \mathbf{u} + \alpha\mathbf{v} > \\ &= < \mathbf{u}, \mathbf{u} > + < \mathbf{u}, \alpha\mathbf{v} > + < \alpha\mathbf{v}, \mathbf{u} > + < \alpha\mathbf{v}, \alpha\mathbf{v} > \\ &= \|\mathbf{u}\|^2 + \overline{\alpha} < \mathbf{u}, \mathbf{v} > + \alpha < \mathbf{v}, \mathbf{u} > + \alpha\overline{\alpha}\|\mathbf{v}\|^2 \end{aligned} \tag{4.11}$$

Since the inner product and $\alpha$ are in general complex scalars, let

$$< \mathbf{u}, \mathbf{v} >= | < \mathbf{u}, \mathbf{v} > |e^{i\theta} \quad \text{and} \quad \alpha = re^{i\theta} \tag{4.12}$$

where $r$ is the modulus and $\theta$ the phase of the complex quantity. Substituting, Eqs. 4.12 into Eq. 4.11,

$$0 \leq \|\mathbf{u}\|^2 + 2r| < \mathbf{u}, \mathbf{v} > | + r^2\|\mathbf{v}\|^2 \equiv f(r) \tag{4.13}$$

$f(r)$ in Eq. 4.13 is a quadratic, in $r$. Since $f(r) \geq 0$, the discriminant $\Delta \leq 0$. When, $f(r) = 0$, the quadratic has two real roots and $f(r) > 0$ corresponds to the situation of two imaginary roots. Hence

$$b^2 - 4ac \leq 0 \quad \rightarrow \quad 4| < \mathbf{u}, \mathbf{v} > |^2 - 4\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \leq 0$$

and

$$| < \mathbf{u}, \mathbf{v} > | \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

which is the Schwarz inequality. There are alternate ways to derive the Schwarz inequality and one such variant is illustrated below.

$$\begin{aligned} 0 &\leq < \mathbf{u} - \alpha\mathbf{v}, \mathbf{u} - \alpha\mathbf{v} > \\ &= \|\mathbf{u}\|^2 - \overline{\alpha} < \mathbf{u}, \mathbf{v} > -\alpha < \mathbf{v}, \mathbf{u} > + \alpha\overline{\alpha}\|\mathbf{v}\|^2 \end{aligned} \tag{4.14}$$

Substituting

$$\alpha = \frac{< \mathbf{u}, \mathbf{v} >}{< \mathbf{v}, \mathbf{v} >} = \frac{< \mathbf{u}, \mathbf{v} >}{\|\mathbf{v}\|^2}$$

in Eq. 4.14, we get

$$0 \leq \|\mathbf{u}\|^2 - \frac{| < \mathbf{u}, \mathbf{v} > |^2}{\|\mathbf{v}\|^2}$$

which yields the Schwarz inequality.

**The Triangular Inequality**

We illustrate how the Schwarz inequality can be used to prove the triangular inequality,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= <\mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}> \\
&= <\mathbf{u},\mathbf{u}> + <\mathbf{v},\mathbf{u}> + <\mathbf{u},\mathbf{v}> + <\mathbf{v},\mathbf{v}> \\
&= \|\mathbf{u}\|^2 + \overline{<\mathbf{u},\mathbf{v}>} + <\mathbf{u},\mathbf{v}> + \|\mathbf{v}\|^2 \\
&= \|\mathbf{u}\|^2 + 2Re <\mathbf{u},\mathbf{v}> + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2|<\mathbf{u},\mathbf{v}>| + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \qquad \text{(Using the Schwartz Inequality)} \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

which yields the triangular inequality,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

## 4.5 The Adjoint Operator

Consider the operator $\mathbf{L}$ on an inner product space X. $\mathbf{L}^*$ is said to be the adjoint of $\mathbf{L}$ if it satisfies the following identity,

$$< \mathbf{L}u, v >=< u, \mathbf{L}^* v > \qquad \forall\, u, v \in X \tag{4.15}$$

The above identity provides a formal route to identifying the adjoint operator $\mathbf{L}^*$. If $\mathbf{L} = \mathbf{L}^*$ then the operator is said to be self-adjoint. The above definition of the adjoint operator is general, and can be used to identify the adjoints for matrix, differential and integral operators without loss of generality. Further, $X$ represents a vector space if $\mathbf{L}$ is a matrix or a function space if $\mathbf{L}$ is either a differential or integral operator. This definition should not be confused with the adjugate or adjoint of a matrix discussed earlier in connection with finding the inverse of the matrix. We illustrate the procedure for finding the adjoint operator starting with matrices.

**Example**: Let $\mathbf{L}$ be the $n \times n$ matrix, $\mathbf{A}$ and $\mathbf{u}$ and $\mathbf{v}$ represent $n$ dimensional vectors.

$$\begin{aligned}
< \mathbf{Au}, \mathbf{v} > &= \sum_i \sum_j a_{ij} u_j \, v_i^* \\
&= \sum_i \sum_j u_j \, a_{ij} v_i^* \\
&= \sum_i \sum_j u_i \, a_{ji} v_j^* \\
&= \sum_i u_i \sum_j a_{ji} v_j^* \\
&= < \mathbf{u}, \sum_j a_{ji}^* v_j > \\
&= < \mathbf{u}, \mathbf{A}^* \mathbf{v} >
\end{aligned}$$

From the last two lines of the above manipulations, it should be clear that the adjoint operator $\mathbf{A}^*$ is simply the Hermitian transpose of $\mathbf{A}$. It the matrix is real symmetric or Hermitian then $\mathbf{A} = \mathbf{A}^*$. Hence symmetric or Hermitian matrices belong to a class of self-adjoint operators that we are already familiar with. In Example 1, we interchanged indices on the 3rd line of the derivation. The reader should be familiar with this index manipulation and derive the definition of the adjoint for an $n \times m$ matrix as an exercise.

**Example 2:** Let $\mathbf{A}$ be matrix with real elements,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

then,

$$\begin{aligned} < \mathbf{Au}, \mathbf{v} > &= \sum_i \sum_j a_{ij} u_j \, v_i \equiv (a_{11}u_1 + a_{12}u_2)v_1 + (a_{21}u_1 + a_{22}u_2)v_2 \\ &= \sum_i \sum_j u_i \, a_{ji} v_j \equiv u_1(a_{11}v_1 + a_{21}v_2) + u_2(a_{12}v_1 + a_{22}v_2) \\ &= < \mathbf{u}, \mathbf{A}^* \mathbf{v} > \end{aligned}$$

where the adjoint,

$$\mathbf{A}^* = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

**Example 3:** Let us consider a specific matrix with complex elements,

$$\mathbf{A} = \begin{pmatrix} i & 0 \\ i & 1 \end{pmatrix}$$

then,

$$\begin{aligned} < \mathbf{Au}, \mathbf{v} > &= iu_1 v_1^* + (iu_1 + u_2)v_2^* \\ &= iu_1 v_1^* + iu_1 v_2^* + u_2 v_2^* \\ &= < \mathbf{u}, \mathbf{A}^* \mathbf{v} > \end{aligned}$$

In this example $\mathbf{A} \neq \mathbf{A}^*$ and hence the matrix is not self-adjoint. $< \mathbf{Au}, \mathbf{v} > = < \mathbf{u}, \mathbf{A}^* \mathbf{v} >$ by definition (Eq. 4.15). However $\mathbf{A} \neq \mathbf{A}^*$. Hence, Eq. 4.15 only provides a prescription for identifying the adjoint operator.

## 4.6 Adjoints for Differential Operators

Consider the differential operator,

$$\mathbf{L}u = \frac{d^2u}{dx^2} + \alpha u(x) = 0 \qquad 0 \leq x \leq 1$$

with the boundary conditions, $u'(0) = 0$ and $u'(1) = 0$. The prime denotes differentiation with respect to $x$. In order to obtain the adjoint operator $\mathbf{L}^*$, we proceed in the following manner.

$$\begin{aligned} < \mathbf{L}u, v > &= \int_0^1 [\frac{d^2u}{dx^2} + \alpha u(x)]v(x)\, dx \\ &= \int_0^1 \frac{d^2u}{dx^2} v(x)\, dx + \int_0^1 \alpha u(x)v(x)\, dx \\ &= [vu' - v'u]_0^1 + \int_0^1 u\, \frac{d^2v}{dx^2}\, dx + \int_0^1 \alpha u(x)v(x)\, dx \end{aligned}$$

The last line is obtained by integrating the term containing $u''(x)$ terms twice by parts. The last step in obtaining the adjoint operator requires incorporating the boundary conditions on $u(x)$. If $B(u, v)$ represents the boundary terms, then

$$\begin{aligned} B(u,v) &= [vu' - v'u]_0^1 \\ &= [v(1)u'(1) - v'(1)u(1) - v(0)u'(0) + v'(0)u(0)] \\ &= [-v'(1)u(1) + v'(0)u'(0)] \qquad \text{since} \qquad u'(0) = 0, u'(1) = 0 \end{aligned}$$

The boundary conditions on $v(x)$ are chosen such that $B(u, v)$ vanishes. This results in $v'(0) = 0$ and $v'(1) = 0$ and

$$\begin{aligned} < \mathbf{L}u, v > &= \int_0^1 u(x)[\frac{d^2v}{dx^2} + \alpha v(x)]\, dx \\ &= < u, \mathbf{L}^*v > \end{aligned}$$

Hence the adjoint operator

$$\mathbf{L}^*v = \frac{d^2v}{dx^2} + \alpha v(x) = 0 \qquad 0 \le x \le 1$$

with the boundary conditions, $v'(0) = 0$ and $v'(1) = 0$. The above prescription formally defines the adjoint operator. Note that both $\mathbf{L}$ and $\mathbf{L}^*$ are defined along with their boundary conditions. The boundary conditions for $\mathbf{L}^*$ were obtained with the requirement that the boundary functional $B(u, v) = 0$. Further, $\mathbf{L} = \mathbf{L}^*$, and the differential operator is said to be self-adjoint. In the case of differential equations, $\mathbf{L} = \mathbf{L}^*$ only when the boundary conditions for $\mathbf{L}$ and the adjoint operator $\mathbf{L}^*$ are identical. This is the case with the example above.

## 4.7 Existence and Uniqueness for $\mathrm{Ax} = \mathrm{b}$ Revisited

We return to the question of existence and uniqueness for linear operators in a more general setting. These theorems are also referred to as the Fredholms alternative theorems and provide a prescription for analyzing the existence and uniqueness conditions for all linear operators. Let us consider the existence and uniqueness conditions for the matrix equation and introduce the concept of the adjoint operator to tackle the existence and uniqueness condition. Consider the matrix equation,

$$\mathbf{Au} = \mathbf{b}$$

1. We first analyze the homogeneous problem,

$$\mathbf{Au} = 0.$$

If $\mathbf{Au} = 0$ has only the trivial solution, then $\mathbf{Au} = \mathbf{b}$ has a unique solution. If $\mathbf{A}$ is an $n \times n$ matrix then this is true for any vector $\mathbf{b}$. However for a $n \times m$ matrix $\mathbf{Au} = \mathbf{b}$ has a unique solution only when the system is solvable. We have examined the proof of the above statements in detail in the previous Chapter.

2. The second part of the theorem concerns the conditions for the solvability (or existence condition) of $\mathbf{Au} = \mathbf{b}$. If $\mathbf{Au} = 0$ has non-trivial solutions we have seen earlier that $\mathbf{Au} = \mathbf{b}$, can have no solution or have an infinity of solutions. In order to determine the conditions for solvability, we examine the homogeneous adjoint problem,

$$\mathbf{A}^*\mathbf{v} = 0 \tag{4.16}$$

where $\mathbf{A}^*$ is the adjoint of $\mathbf{A}$. The theorem states that $\mathbf{Au} = \mathbf{b}$ has a solution if and only if

$$< \mathbf{b}, \mathbf{v} >= 0 \qquad \forall\, \mathbf{v}\; s.t.\; \mathbf{A}^*\mathbf{v} = 0 \tag{4.17}$$

The above condition provides the solvability or existence condition for the inhomogeneous problem. The statement in Eq. 4.17 is equivalent to stating that the rhs vector $\mathbf{b}$ is orthogonal to the null space of the adjoint operator, $\mathbf{A}^*$ since $\mathbf{v}$ satisfies, Eq. 4.16. To show that when $\mathbf{u}$ is a solution to $\mathbf{Au} = \mathbf{b}$ then $< \mathbf{b}, \mathbf{v} >= 0$, where $\mathbf{v}$ satisfies $\mathbf{A}^*\mathbf{v} = 0$.

Proof: Since

$$\mathbf{Au} = \mathbf{b}$$

it follows that,

$$< \mathbf{Au}, \mathbf{v} >=< \mathbf{b}, \mathbf{v} > \tag{4.18}$$

Now

$$< \mathbf{Au}, \mathbf{v} >=< \mathbf{u}, \mathbf{A}^*\mathbf{v} >= 0 \tag{4.19}$$

Hence

$$< \mathbf{b}, \mathbf{v} >= 0$$

To complete the proof we need to show that, if $< \mathbf{b}, \mathbf{v} >= 0$ then $< \mathbf{Au} = \mathbf{b} >$ has a solution, i.e, $\mathbf{b}$ lies in the range of the operator. We do not pursue this here. The theorem can be used to verify the solvability condition when $\mathbf{A}$ is nonsingular. When $\mathbf{A}$ is nonsingular then the only solution to $\mathbf{A}^*\mathbf{v} = 0$ is the trivial solution. Hence $< \mathbf{b}, \mathbf{v} >= 0$ for any $\mathbf{b}$ and $\mathbf{Ax} = \mathbf{b}$ is therefore solvable for all $\mathbf{b}$. Although we have proved the alternative theorems developed above using $\mathbf{A}$ as the linear operator, the theorems are true for linear operator in general. We will use these alternative theorems to study the existence and uniqueness conditions for some differential operators later in the text.

**Example 4**: In this example we use the solvability condition of the alternative theorem to identify the range space for the set of linear equations,

$$\begin{aligned} x_1 + x_2 + x_3 &= b_1 \\ 2x_1 - x_2 + x_3 &= b_2 \\ x_1 - 2x_2 &= b_3 \end{aligned}$$

We first identify the null space vectors for $\mathbf{A}^*$ using elementary row operations

$$\mathbf{A}^* = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & -2 \\ 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The basis for the null space of $\mathbf{A}^*$,

$$\mathbf{v} = \alpha \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

The solvability condition states that $\mathbf{Ax} = \mathbf{b}$ is solvable if and only if $< \mathbf{b}, \mathbf{v} >= 0$. This results in $b_1 - b_2 + b_3 = 0$ which yields the following basis for the range space of $\mathbf{A}$,

$$\mathbf{b} = \alpha \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

The range space vectors can also be obtained using the rank equivalence criterion. The reader should obtain and compare the range space vectors using the rank criterion.

We end this Chapter with by using the Fredholms alternative theorem to prove the following theorem concerning the dimensions of $\mathcal{N}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A})$.

**Theorem:** For a general $m \times n$ matrix $\mathbf{A}$

$$\dim \mathcal{N}(\mathbf{A}) + \dim \mathcal{R}(\mathbf{A}) = n \tag{4.20}$$

where $\dim \mathcal{N}(\mathbf{A}) = n - r$ and the $\dim \mathcal{R}(\mathbf{A}) = r$

Case 1: Let $m = n$. The dimension of $\mathcal{N}(\mathbf{A}) = n - r$. Since the rank of $\mathbf{A}^*$ is the same as the rank of $\mathbf{A}$, the dimension of $\mathcal{N}(\mathbf{A}^*) = n - r$. In order to determine the dimension of $\mathcal{R}(\mathbf{A})$ we utilize the solvability condition based on the Fredholms alternative theorem. $\mathbf{Ax} = \mathbf{b}$ is solvable if and only if

$$< \mathbf{b}, \mathbf{v}_i >= 0 \qquad i = 1, 2, \ldots n - r \tag{4.21}$$

where $\mathbf{v}_i \in \mathcal{N}(\mathbf{A}^*)$ i.e. $\mathbf{A}^*\mathbf{v}_i = 0$. Since $\mathbf{b}$ is a column vector with $n$ unknowns, Eq. 4.21 provides $n - r$ equations. $r$ unknowns can be chosen independently, resulting in $\dim \mathcal{R}(\mathbf{A}) = r$. Hence Eq. 4.20 is true.

Case 2: Let $\mathbf{A}$ be an $m \times n$ matrix of rank $r$. The dimension of $\mathcal{N}(\mathbf{A}) = n - r$. Since the rank of $\mathbf{A}^*$ is the same as the rank of $\mathbf{A}$ and $\mathbf{A}^*$ is an $n \times m$ matrix, the dimension of $\mathcal{N}(\mathbf{A}^*) = m - r$. Solvability conditions results in,

$$< \mathbf{b}, \mathbf{v}_i >= 0 \qquad i = 1, 2, \ldots m - r$$

where $\mathbf{v}_i \in \mathcal{N}(\mathbf{A}^*)$ i.e. $\mathbf{A}^*\mathbf{v}_i = 0$. Since $\mathbf{b}$ is a column vector with $m$ unknowns, Eq. 4.7 provides $m - r$ equations. $r$ equations can be chosen independently resulting in $\dim \mathcal{R}(\mathbf{A}) = r$. Hence Eq. 4.20 is true. The proof for Case 2, is valid for both $m < n$ or $m > n$. In either case rank, $r \leq min(m, n)$

## Problems

1. **Solvability Conditions**

   Use the Fredholm's alternative theorem to determine the solvability conditions (existence) for the following sets of linear equations, by checking if the right hand side vector $\mathbf{b}$ is orthogonal to the null space of $\mathbf{A}^*$. If the system is solvable, comment on the uniqueness of the solution.

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 3 \\ 2x_1 + x_2 + 6x_3 &= 2 \\ x_2 + 2x_2 + 4x_3 &= -1 \end{aligned}$$

$$\begin{aligned} x_1 + 2x_2 + x_3 + 2x_4 - 3x_5 &= 2 \\ 3x_1 + 6x_2 + 4x_3 - x_4 + 2x_5 &= -1 \\ 4x_1 + 8x_2 + 5x_3 + x_4 - x_5 &= 1 \\ -2x_1 - 4x_2 - 3x_3 + 3x_4 - 5x_5 &= 3 \end{aligned}$$

$$\begin{aligned} x_1 - x_2 + 3x_3 + 2x_4 &= 2 \\ 3x_1 + x_2 - x_3 + x_4 &= -3 \\ -x_1 - 3x_2 + 7x_3 + 3x_4 &= 7 \end{aligned}$$

2. **Gram Schmidt Orthogonalization**

   Find the eigenvalues and eigenvectors of

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

   (a) Show that the eigenvectors form a linearly independent set.

   (b) Using the Gramd-Schmidt process construct an orthonormal set of eigenvectors.

3. **Gram Schmidt Orthogonalization**

   Consider the following set of 5 vectors

$$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

(a) Using the above vectors construct a subset containing the maximum number of linearly independent vectors.

(b) Using the set obtained in part (a) above construct an orthonormal set of vectors using Gram-Schmidt orthogonalization.

4. **Orthonormal Functions**

Show that the following functions

$$\phi_n(x) = exp(2\pi i\, nx) \qquad n = 0, \pm 1, \pm 2 \ldots, \qquad 0 \leq x \leq 1$$

where $i = \sqrt{-1}$ form an orthonormal set.

5. **Fourier Series Representation of Functions**

Consider a piecewise continuous function $f(x)$ defined on the interval $[-c, c]$ with period $2c$. The function can be represented as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n cos \frac{n\pi x}{c} + b_n sin \frac{n\pi x}{c} \right)$$

(a) Determine expressions for the coefficients $a_n$ and $b_n$.

(b) Simplify the series expansions for odd functions $f(x)$ and even functions $f(x)$.

(c) For a function

$$f(x) = \begin{cases} -\pi/2 & -\pi < x < 0 \\ \pi/2 & 0 < x < \pi \end{cases}$$

and $f(0) = 0$, evaluate the Fourier series representation.

(d) If $S_M(x)$ is the value of the series with $M$ terms in the summation, then plot $S_M(x)$ for the series obtained in part (c) for different values of $M$. What can you conclude about the series representation for $f(x)$ ?

(e) It can be shown that

$$\lim_{M \to \infty} S_M \left( \frac{\pi}{2M} \right) = \int_0^{\pi} \frac{sinx}{x} dx$$

Use this result to check your limiting value of the summation that you compute.

6. **Fourier Series Solutions**

The Fourier series solution to the temperature $T(x,t)$ in a time dependent 1D heat conduction problem is

$$T(x,t) = \sum_{n=1}^{\infty} a_n exp(-\alpha n^2 \pi^2 t/c^2) sin\frac{n\pi x}{c}$$

where $\alpha$ is the thermal diffusivity.

(a) Using the initial condition $T(x, t = 0) = f(x)$ obtain an expression for the coefficients $a_n$ in the expansion.

(b) You will need to evaluate the following integral

$$\int_0^c sin\frac{n\pi x}{c}\, sin\frac{m\pi x}{c}\, dx$$

for $n = m$ and $n \neq m$.

(c) If the initial condition $f(x) = x$, carry out the integrations and obtain an expression for the coefficients $a_n$ in the expansion

7. **Fourier Series Solutions**

The Fourier series solution to the temperature $T(x,y,t)$ in a time dependent 2D heat conduction problem on a rectangle with sides of length $a$ and $b$ can be expressed as

$$T(x,y,t) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{nm} exp[-(n^2\pi^2/a^2 + m^2\pi^2/b^2)t] sin\frac{n\pi x}{a} sin\frac{m\pi y}{b}$$

(a) Using the initial condition $T(x, y, t = 0) = f(x, y)$ obtain a general expression for the coefficients $a_{nm}$ in the expansion. You will need to evaluate the following integral

$$\int_0^a sin\frac{n\pi x}{a}\, sin\frac{m\pi x}{a}\, dx$$

for $n = m$ and $n \neq m$.

(b) If the initial condition $f(x) = xy$, carry out the integrations and obtain an expression for the coefficients $a_{nm}$ in the expansion

8. **Orthogonal Functions**

Consider the following ode

$$\frac{d^2u_n}{dx^2} + \lambda_n^2 u_n = 0$$

with the boundary conditions $u_n'(x = 0) = u_n(x = 1) = 0$.

(a) Obtain the general solution to this equation.

(b) Obtain non-trivial solutions of the form $u_1(\lambda_1 x), u_2(\lambda_2 x) \ldots u_n(\lambda_n x)$ for the above boundary conditions. What are the values of $\lambda_n$?

(c) Verify that the solutions $u_1(x), u_2(x) \ldots$ form an orthogonal set. Construct an orthonormal set of functions.

(d) Evaluate the following integrals

$$\int_0^1 sinn\pi x\, sinm\pi x\, dx \qquad \int_0^1 cosn\pi x\, cosm\pi x\, dx$$

for $n, m = 0, 1 \ldots$. While evaluating the integrals you will have to treat the cases $n \neq m$ and $n = m$ separately.

Note: The above ode arises while solving partial differential equations with the separation of variables method where the functions $u_n(x)$ are known as the eigenfunctions and $\lambda_n$ are the eigenvalues.

9. **Gram-Schmidt Orthogonalization**

Consider the functions

$$\phi_n(x) = exp(-x/2)x^n \quad 0 \leq x \leq \infty$$

(a) Using Gram-Schmidt orthogonalization construct an orthonormal basis $\psi_n(x)$ for $n$ = 0, 1 and 2.

(b) Show that the orthonormal basis forms a linearly independent set.

10. **Series Expansions**

Consider the following expansion in a basis

$$f(x) = \sum_{n=1}^{\infty} a_n \phi_n(x)$$

(a) If the weighted inner product,

$$\langle \phi_n \phi_m \rangle_{w(x)} = \int_a^b \phi_n(x)\phi_m(x)w(x) = \delta_{nm}$$

obtain an expression for the coefficients $a_n$.

(b) Consider now the finite series representation of $f(x)$

$$f(x) \approx \sum_{n=1}^{M} c_n \phi_n(x).$$

Obtain the coefficients $c_n$ by minimizing the least square error

$$\int_a^b \left[ f(x) - \sum_{n=1}^{M} c_n \phi_n(x) \right]^2 w(x)dx.$$

(c) Comment on the value of the coefficients $a_n$ and $c_n$.

11. Prove the following:

(a)

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{4.22}$$

(b)

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \tag{4.23}$$

(c)

$$\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \leq \|\mathbf{x} - \mathbf{y}\| \tag{4.24}$$

Give a geometric interpretation for a) and b)

12. Consider the Bessel's inequality

$$\sum_{i=1}^{M} | < \mathbf{e_i}, \mathbf{x} > | \leq \|\mathbf{x}\|^2 \tag{4.25}$$

where $\mathbf{e_i}$ denote the orthonormal basis in the M-dimensional vector space. If, N denotes the dimension of vector space into which $\mathbf{x}$ can be decomposed show that the equality sign holds if $M = N$. What is the condition under which inequality sign holds ?

13. Consider the Schwartz inequality

$$| < \mathbf{x}, \mathbf{y} > | \leq \|\mathbf{x}\|\|\mathbf{y}\| \tag{4.26}$$

For non-zero $||\mathbf{x}||$ and $||\mathbf{y}||$ show that the equality holds if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent. Interpret this geometrically.

14. Use the inner product to verify the following identities

(a)

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) \tag{4.27}$$

(b)

$$\|\mathbf{z} - \mathbf{x}\|^2 + \|\mathbf{z} - \mathbf{y}\|^2 = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 + 2\|\mathbf{z} - \frac{1}{2}(\mathbf{x} + \mathbf{y})\|^2 \tag{4.28}$$

# Chapter 5

# Eigenvalues and Eigenvectors

**Definition**: A complex number $\lambda$ is an eigenvalue of $\mathbf{A}$ if there exists a non-zero vector $\mathbf{x}$ called the eigenvector such that

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{5.1}$$

Eq. 5.1 can be rewritten as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0 \tag{5.2}$$

From Eq. 5.2, eigenvectors $\mathbf{x}$ belong to the null space of $(\mathbf{A} - \lambda\mathbf{I})$ and $\lambda$'s are scalars which result in a zero determinant for, $\mathbf{A} - \lambda\mathbf{I}$. The null space of $\mathbf{A} - \lambda\mathbf{I}$ is also referred to as the eigenspace corresponding to the eigenvalue $\lambda$. The eigenvectors corresponding to a particular eigenvalue form a basis for the eigenspace. In this Chapter our primary focus will be to answer the following questions. Given an $n \times n$ matrix $\mathbf{A}$, can we always obtain $n$ linearly independent eigenvectors? Under what conditions do these eigenvectors form an orthonormal set? Can these eigenvectors be used to solve nonhomogeneous problems of the kind $\mathbf{A}\mathbf{x} = \mathbf{b}$ and initial value problems of the following form,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{b}(t)$$

Given an $n \times n$ matrix, $\mathbf{A}$ the eigenvalues $\lambda_i$'s $i = 1 \ldots n$, are obtained by solving the characteristic equation

$$|\mathbf{A} - \lambda\mathbf{I}| \equiv f(\lambda) = 0 \tag{5.3}$$

The *algebraic multiplicity* for a given eigenvalue $\lambda_i$ is the number of times the root $\lambda_i$ is repeated. The **geometric multiplicity** of $\lambda_i$ is the dimension of the vector space spanned by the

eigenvectors corresponding to the eigenvalue $\lambda_i$. Equivalently the geometric multiplicity corresponding to $\lambda_i$ is nothing but the dimension of the null space of $\mathbf{A} - \lambda_i \mathbf{I}$. The geometric multiplicity cannot exceed the algebraic multiplicity. We will see that it is desirable to have matrices where the geometric multiplicity is equivalent to the algebraic multiplicity.

Given a matrix $\mathbf{A}$ and its corresponding eigenvalues we are interested in the properties of the eigenvectors. If the eigenvectors are to be used as a basis set, they would have to be linearly independent. Further, as we saw in the last Chapter it would be desirable to form a basis with an orthogonal set. Theorem 1, is concerned with the linear dependence of eigenvectors and Theorem 2, addresses the issue of orthogonality between eigenvectors.

**Theorem 1**: Eigenvectors corresponding to distinct eigenvalues are linearly independent.

**Proof**: Let the matrix $\mathbf{A}$ have eigenvalues, $\lambda_i$, $i = 1, \ldots n$. Hence

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i \qquad i = 1 \ldots, n \tag{5.4}$$

If the eigenvectors form a linearly independent set, then the only solution to

$$\sum_i^n c_i \mathbf{x}_i = 0 \tag{5.5}$$

is when $c_i$'s are identically zero. Premultiplying Eq. 5.5 sequentially by $\mathbf{A}$, we can generate the following set of linear $n$ algebraic equations,

$$\begin{aligned}
\sum_i^n c_i \mathbf{x}_i &= 0 \\
\sum_i^n c_i \lambda_i \mathbf{x}_i &= 0 \\
\sum_i^n c_i \lambda_i^2 \mathbf{x}_i &= 0 \\
&\vdots \\
\sum_i^n c_i \lambda_i^{n-1} \mathbf{x}_i &= 0
\end{aligned} \tag{5.6}$$

which can be written in matrix vector notation as,

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1} \end{pmatrix} \begin{pmatrix} c_1\mathbf{x}_1 \\ c_2\mathbf{x}_2 \\ \vdots \\ c_n\mathbf{x}_n \end{pmatrix} = 0 \tag{5.7}$$

Since the eigenvalues are distinct the above matrix is non-singular and the determinant is nonzero. The above matrix is also known as the Vandermonde matrix and it can be shown that the determinant is

$$\prod_{i,j=1}^{n} (\lambda_j - \lambda_i) \neq 0 \qquad (j \neq i)$$

Since Eq. 5.7 represents a set of homogeneous equations, the only solution is the trivial solution. Further since $\mathbf{x}_i$ are the eigenvectors they are non-zero by definition and $c_i$'s are identically zero.

**Theorem 2**: If $\mathbf{L}$ is a self-adjoint operator, i.e. $\mathbf{L} = \mathbf{L}^*$, then the eigenvalues and eigenvectors of $\mathbf{L}$ satisfy the following properties.

1. The eigenvalues of $\mathbf{L}$ are real.
2. Eigenvectors corresponding to distinct eigenvalues are orthogonal.

**Proof 1**:

$$\mathbf{L}u = \lambda u$$

Taking inner products with $u$

$$< \mathbf{L}u, u >= \lambda < u, u > \tag{5.8}$$

Using the definition of the adjoint operator

$$\begin{aligned} < \mathbf{L}u, u > &= < u, \mathbf{L}^*u > \\ &= < u, \mathbf{L}u > \qquad \text{(Since } \mathbf{L} = \mathbf{L}^*\text{)} \\ &= \overline{\lambda} < u, u > \end{aligned} \tag{5.9}$$

From Eqs. 5.8 and 5.9, $\lambda = \overline{\lambda}$. This is only possible when $\lambda$ is real.

**Proof 2:** Let

$$\mathbf{L}u = \lambda_u u \qquad \text{and} \qquad \mathbf{L}v = \lambda_v v \tag{5.10}$$

Taking inner products of the first equation in Eq. 5.10 with $v$,

$$< \mathbf{L}u, v >= \lambda_u < u, v > \tag{5.11}$$

Using the definition of the adjoint operator,

$$\begin{aligned} < \mathbf{L}u, v > &= < u, \mathbf{L}^* v > \\ &= \overline{\lambda_v} < u, v > \\ &= \lambda_v < u, v > \qquad (\text{Since } \lambda \text{ is real}) \end{aligned} \tag{5.12}$$

Equating Eq. 5.11 with Eq. 5.12 we get,

$$(\lambda_u - \lambda_v) < u, v >= 0 \tag{5.13}$$

Since $\lambda_u \neq \lambda_v$ and $u$ and $v$ are non-zero by definition, Eq. 5.13 implies that $< u, v >= 0$, i.e $u$ is orthogonal to $v$.

Although we are presently occupied with matrix operators, the above proof is valid for self-adjoint operators in general. The proof also illustrates the utility of using innner products in proving the theorem. We illustrate the implications of Theorem 1 and Theorem 2 with some examples. We first consider some illustrative examples for nonsymmetric matrices. As an exercise, the reader should obtain the eigenvalues and eigenvectors for the examples given below.

**Example 1**: Nonsymmetric matrix, distinct eigenvalues

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & = & 3 \\ \lambda_2 & = & -1 \end{pmatrix} \qquad \mathbf{x}^{(1)} = \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad \mathbf{x}^{(2)} = \beta \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

The superscript $i$ on the eigenvector corresponds to $i$th eigenvalue.

**Example 2**: Nonsymmetric matrix, multiple eigenvalues

$$\mathbf{A} = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & = & 2 \\ \lambda_2 & = & 2 \end{pmatrix} \qquad \mathbf{x}^{(1)} = \alpha \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The first example illustrates that when a nonsymmetric matrix has distinct eigenvalues, it is possible to obtain two distinct eigenvectors. By Theorem 1, these eigenvectors are linearly

independent. The second example illustrates a situation where a nonsymmetric matrix with multiple eigenvalues has only one eigenvector. This is an example where the geometric multiplicity is less than the algebraic multiplicity. For a given $\lambda_i$ with multiplicity $m$, this situation occurs for non-symmetric matrices when dim $\mathcal{N}(\mathbf{A} - \lambda I) < m$. This is symptomatic of situations where the algebraic multiplicty is greater than unity for nonsymmetric matrices. In the next two examples we will consider Hermitian matrices.

**Example 3**: Symmetric matrix, distinct eigenvalues

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & = & 3 \\ \lambda_2 & = & -1 \end{pmatrix} \qquad \mathbf{x}^{(1)} = \alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{x}^{(2)} = \beta \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The eigenvectors are not only linearly independent, but are also orthogonal by Theorem 2. The orthonormal set is obtained by dividing each eigenvector by its norm. The orthonormal set is,

$$\mathbf{x}^{(1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{x}^{(2)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

**Example 4**: Symmetric matrix with multiple eigenvalues

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & = & 2 \\ \lambda_2 & = & 2 \end{pmatrix} \qquad \mathbf{x}^{(1)} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{x}^{(2)} = \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

In this example the matrix has one eigenvalue of multiplicity 2. However unlike the situation in Example 2, here we are able to obtain two distinct eigenvectors which constitute an orthonormal set. We can state the following theorem for Hermitian matrices.

**Theorem:** A Hermitian matrix of order $n$ has $n$ linearly independent eigenvectors and these form an orthonormal set.

Consider a Hermitian matrix with distinct eigenvalues. From Theorem 2, it is clear that the eigenvectors corresponding to distinct eigenvalues are orthogonal and the eigenvectors form an orthonormal set. If the matrix has multiple eigenvalues then the orthornormal set is constructed by using Gram-Schmidt orthogonalization. Consider an $n \times n$ Hermitian matrix with $k$ distinct eigenvalues, $\lambda_1, \lambda_2, \ldots \lambda_k$. Let the multiplicity of the $k+1$th eigenvalue $\lambda_{k+1}$ be $m$. For the first $k$ set of distinct eigenvalues there are $k$ eigenvectors which constitute an orthogonal set (Theorem 2, part 2). Each of these $k$ eigenvectors are orthogonal to the remaining $m$ eigenvectors corresponding the eigenvalue $\lambda_{k+1}$ which has multiplicity $m$ (Theorem 2, part 2). The missing piece is the orthogonality between the $m$ eigenvectors corresponding to the

repeated eigenvalue $\lambda_{k+1}$. Since these are eigenvectors that belong to the same eigenvalue, Theorem 2, does not apply. However we can use Gram-Schmidt orthogonalization to construct an orthonormal set of these $m$ eigenvectors. With this the construction is complete and we have a set of $n$ orthonormal eigenvectors. We note that the Gram-Schmidt orthogonalization is essentially a process of taking linear combinations of vectors and we need to shown that the new vectors are still eigenvectors having the same eigenvalue. The following Lemma concerns this point.

**Lemma**: Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_m$ be $m$ eigenvectors corresponding the eigenvalue $\lambda$. Hence

$$\mathbf{A}\mathbf{x}_i = \lambda \mathbf{x}_i \qquad i = 1, \ldots m$$

The eigenvalue of the eigenvector constructed by taking linear combinations of the $m$ eigenvectors is also $\lambda$.

Proof: Let $\mathbf{y}$ be the eigenvector obtained by taking a linear combination of $m$ eigenvectors,

$$\mathbf{y} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i$$

Now

$$\begin{aligned} \mathbf{A}\mathbf{y} &= \mathbf{A}(\sum_{i=1}^{m} \alpha_i \mathbf{x}_i) \\ &= \sum_{i=1}^{m} \alpha_i \mathbf{A}\mathbf{x}_i \\ &= \sum_{i=1}^{m} \alpha_i \lambda \mathbf{x}_i \\ &= \lambda \sum_{i=1}^{m} \alpha_i \mathbf{x}_i \\ &= \lambda \mathbf{y} \end{aligned} \tag{5.14}$$

Hence $\mathbf{y}$ is an eigenvector with eigenvalue $\lambda$. We note that this proof is true for linear combinations which involve any subset of the $m$ eigenvectors.

Finally we state (without proof) that given a Hermitian matrix, the algebraic multiplicity always equals the geometric multiplicity. This implies that for an eigenvalue with mulitplicity

$m$,

$$\dim \mathcal{N}(\mathbf{A} - \lambda_i I) = m$$

Hence for a Hermitian matrix we are ensured of finding all the eigenvectors regardless of the mulitplicities in the eigenvalues. This concludes the proof for Theorem 3.

## 5.1 Eigenvectors as Basis Sets

Consider the matrix equation

$$\mathbf{A}\mathbf{u} = \mathbf{b} \tag{5.15}$$

We will assume that the matrix posesses $n$ eigenvectors. Further let us assume that the determinant of $\mathbf{A}$ is non-zero i.e $\lambda_i \neq 0$ for $i \ldots n$. Let

$$\mathbf{u} = \sum_i^n c_i \mathbf{x}_i \tag{5.16}$$

Substitute Eq. 5.16 into Eq. 5.15,

$$\begin{aligned} \mathbf{A}\sum_i^n c_i \mathbf{x}_i &= \mathbf{b} \\ \sum_i^n c_i \mathbf{A}\mathbf{x}_i &= \mathbf{b} \\ \sum_i^n c_i \lambda_i \mathbf{x}_i &= \mathbf{b} \end{aligned}$$

Taking inner products with $\mathbf{x}_j$

$$\sum_i^n c_i < \lambda_i \mathbf{x}_i, \mathbf{x}_j > \;=\; < \mathbf{b}, \mathbf{x}_j > \qquad j = 1 \ldots n \tag{5.17}$$

The above manipulations results in a set of $n$ linear algebraic equations which can be compactly represented as

$$\mathbf{M}\mathbf{c} = \mathbf{f}$$

where the elements of $\mathbf{M}$, $m_{ij} =< \lambda_j \mathbf{x}_j, \mathbf{x}_i >$ and $f_i =< \mathbf{b}, \mathbf{x}_i >$. A solution of Eq. 5.1, yields the coefficients in the expansion. If the eigenvectors form an orthonormal set, as would be the

case if $\mathbf{A}$ were Hermitian then the coefficients can be obtained analytically,

$$c_i = \frac{< \mathbf{b}, \mathbf{x}_i >}{\lambda_i}$$

and the solution can be expressed as,

$$\mathbf{u} = \sum_{i=1}^{n} \frac{< \mathbf{b}, \mathbf{x}_i >}{\lambda_i} \mathbf{x}_i \tag{5.18}$$

The above method is general and can also be used when the det $\mathbf{A}$= 0. We leave this as an exercise. The above solution illustrates the utilility of the eigenvectors as a basis while seeking solutions to matrix equations. In the absence of an orthonormal set of eigenvectors, obtaining the coefficients involves solving a set of linear equations. If the eigenvectors form an orthonormal set, as is the case with a Hermitian matrix $\mathbf{A}$, then the solution is greatly simplified.

## 5.2 Similarity Transforms

Very often equations involving matrices can be conveniently treated using suitable transformations. Clearly a transformation that preserves the eigenvalues of the matrix will preserve the underlying physics of the problem. One such transformation is the similarity transform. In this section we introduce the similarity transform and illustrate its utility for matrix diagonalization, matrix algebras and solutions of IVPs.

**Definition:** If there exists a non-singular matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B}$$

then, $\mathbf{B}$ is said to be similar to $\mathbf{A}$. Similar matrices have the same eigenvalues.

**Theorem:** Similar matrices have the same eigenvalues. If $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B}$ then $\mathbf{A}$ is similar to $\mathbf{B}$ and both $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalue.

**Proof:** Let the eigenvalue of $\mathbf{A}$ be $\lambda$.

$$\begin{aligned} \mathbf{B} &= \mathbf{P}^{-1}\mathbf{A}\mathbf{P} \\ \mathbf{B}\mathbf{P}^{-1} &= \mathbf{P}^{-1}\mathbf{A} \\ \mathbf{B}\mathbf{P}^{-1}\mathbf{x} &= \mathbf{P}^{-1}\mathbf{A}\mathbf{x} \\ &= \mathbf{P}^{-1}\lambda\mathbf{x} \\ &= \lambda\mathbf{P}^{-1}\mathbf{x} \end{aligned} \tag{5.19}$$

If $\mathbf{P}^{-1}\mathbf{x} = \mathbf{y}$, the the last line of the above equation implies,

$$\mathbf{B}\mathbf{y} = \lambda\mathbf{y}$$

Hence $\lambda$ is an eigenvalue of $\mathbf{B}$ with eigenvector $\mathbf{P}^{-1}\mathbf{x}$.

### 5.2.1 Diagonalization of $\mathbf{A}$

If a matrix $\mathbf{A}$ has $n$ linearly independent eigenvectors, then

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \Lambda$$

$\mathbf{P}$ is a nonsingular matrix whose columns are made up of the eigenvectors of $\mathbf{A}$ and $\Lambda$ is a diagonal matrix with $\lambda$'s on the diagonal.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0\dots & 0 \\ 0 & 0 & \lambda_3\dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

We next show that the matrix $\mathbf{P}$ made up of the eigenvectors, reduces $\mathbf{A}$ to a diagonal matrix under a similirity transformation. Let $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i \quad i = 1, \dots n$

$$\begin{aligned} \mathbf{A}\mathbf{P} &= \mathbf{A}[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \\ &= [\mathbf{A}\mathbf{x}_1, \mathbf{A}\mathbf{x}_2, \dots, \mathbf{A}\mathbf{x}_n] \\ &= [\lambda_1\mathbf{x}_1, \lambda_2\mathbf{x}_2, \dots, \lambda_n\mathbf{x}_n] \\ &= \mathbf{P}\Lambda \end{aligned} \tag{5.20}$$

In general a square matrix can be reduced to a diagonal matrix if and only if it possesses $n$ linearly independent eigenvectors. This is always possible for Hermitian matrices.

### 5.2.2 Using similarity transforms

Similarity transforms can be used to perform matrix algebras in a convenienent manner as illustrated below,

1. Powers of Matrices

$$\begin{aligned} \mathbf{P}^{-1}\mathbf{A}\mathbf{P} &= \Lambda \\ \mathbf{A} &= \mathbf{P}\Lambda\mathbf{P}^{-1} \\ \mathbf{A}^n &= (\mathbf{P}\Lambda\mathbf{P}^{-1})\,(\mathbf{P}\Lambda\mathbf{P}^{-1})\ldots(\mathbf{P}\Lambda\mathbf{P}^{-1}) \\ &= \mathbf{P}\Lambda^n\mathbf{P}^{-1} \end{aligned} \tag{5.21}$$

2. Inverse

$$\begin{aligned} \mathbf{A}^{-1} &= (\mathbf{P}\Lambda\mathbf{P}^{-1})^{-1} \\ &= (\mathbf{P}\Lambda^{-1}\mathbf{P}^{-1}) \end{aligned} \tag{5.22}$$

3. Matrix polynomials

$$\begin{aligned} f(\mathbf{A}) &= a_0\mathbf{A}^m + a_1\mathbf{A}^{m-1}\ldots a_m\mathbf{I} \\ &\quad (\text{Using, } \mathbf{A}^n = \mathbf{P}\Lambda^n\mathbf{P}^{-1}) \\ &= a_0\mathbf{P}\Lambda^m\mathbf{P}^{-1} + a_1\mathbf{P}\Lambda^{m-1}\mathbf{P}^{-1} + \ldots a_m\mathbf{P}\mathbf{P}^{-1} \\ &= \mathbf{P}(a_0\Lambda^m + a_1\Lambda^{m-1} + \ldots, + a_m\mathbf{I})\mathbf{P}^{-1} \\ &= \mathbf{P}f(\lambda)\mathbf{P}^{-1} \end{aligned} \tag{5.23}$$

In all the above manipulations, the algebra is reduced to taking powers of the diagonal matrix $\Lambda$. To complete the solution, one also requires a knowledge of $\mathbf{P}^{-1}$. Under certain conditions $\mathbf{P}^{-1}$ is easily deduced from $\mathbf{P}$. We discuss this next.

## 5.3 Unitary and orthogonal matrices

**Definition**: $\mathbf{P}$ is said to be a unitary matrix if,

$$\mathbf{P}^*\mathbf{P} = \mathbf{P}\mathbf{P}^* = \mathbf{I},$$

which implies that $\mathbf{P}^{-1} = \mathbf{P}^*$. As defined earlier, $\mathbf{P}^*$ is the complex conjugate transpose of $\mathbf{P}$. If $\mathbf{P}$ consists of real elements then,

$$\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I},$$

which implies that $\mathbf{P}^{-1} = \mathbf{P}^T$. Then $\mathbf{P}$ is said to be orthogonal. Further, if

$$\mathbf{P}^*\mathbf{P} = \mathbf{P}\mathbf{P}^*$$

then the matrix $\mathbf{P}$ is said to be normal. Normal matrices provides a broader classification for matrices which includes both unitary and orthogonal matrices. Other examples of normal matrices are, Hermitian, skew Hermitian and diagonal matrices.

**Theorem**: If $\mathbf{A}$ is a Hermitian matrix then the matrix $\mathbf{P}$ whose columns are made up of the eigenvectors of $\mathbf{A}$ is a unitary matrix.

**Proof**: Since $\mathbf{P}$ is made up of the eigenvectors of $\mathbf{A}$,

$$\mathbf{P} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \ldots \mathbf{x}_n]$$

Then

$$\mathbf{P}^*\mathbf{P} = \begin{pmatrix} \mathbf{x}_1^\dagger\mathbf{x}_1 & \mathbf{x}_1^\dagger\mathbf{x}_2 & \ldots, & \mathbf{x}_1^\dagger\mathbf{x}_n \\ \mathbf{x}_2^\dagger\mathbf{x}_1 & \mathbf{x}_2^\dagger\mathbf{x}_2 & \ldots, & \mathbf{x}_2^\dagger\mathbf{x}_n \\ \vdots & \vdots & \ldots, & \vdots \\ \mathbf{x}_n^\dagger\mathbf{x}_1 & \mathbf{x}_n^\dagger\mathbf{x}_2 & \ldots, & \mathbf{x}_n^\dagger\mathbf{x}_n \end{pmatrix} \tag{5.24}$$

We have introduced some new notation for clarity. In the above equations, $\mathbf{x}_i^\dagger$ represents the complex conjugate transpose of the column vector $\mathbf{x}_i$. Hence,

$$\mathbf{x}_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{then} \qquad \mathbf{x}_i^\dagger = \left(\overline{x_1}, \overline{x_1}, \ldots, \overline{x_n}\right)$$

Noting, that the matrix elements in Eq. 5.24 are inner products between the eigenvectors which form an orthonormal set ($\mathbf{A} = \mathbf{A}^*$).

$$\mathbf{x}_i^\dagger \mathbf{x}_j = < \overline{\mathbf{x}_i, \mathbf{x}_j} > = \delta_{ij}$$

Hence

$$\mathbf{P}^*\mathbf{P} = \mathbf{I}$$

In a similar manner one can show that $\mathbf{P}\mathbf{P}^* = \mathbf{I}$. Therefore, $\mathbf{P}$ is unitary. The proof for orthogonal matrices follows along similar lines, with $\mathbf{x}_i^\dagger$ replaced with $\mathbf{x}_i^T$.

**Example 5:** Consider the non-symmetric matrix from Example 1. Using the two linearly independent eigenvectors to construct matrix $\mathbf{P}$,

$$\mathbf{P} = \begin{pmatrix} 1 & -1 \\ 2 & 2 \end{pmatrix}, \quad \mathbf{P}^{-1} = \frac{1}{4}\begin{pmatrix} 2 & 1 \\ -2 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$$

Note that the order in which the eigenvalues appear in the diagonal matrix is dependent on how the eigenvectors are ordered in the matrix $\mathbf{P}$.

**Example 6:** Consider the symmetric matrix from Example 3.

$$\mathbf{P} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{P}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$$

If $\mathbf{P}$ is constructed using the orthonormal eigenvectors of $\mathbf{A}$ then,

$$\mathbf{P} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \quad \text{and} \quad \mathbf{P}^{-1} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Here $\mathbf{P}^{-1} = \mathbf{P}^T$ and $\mathbf{P}$ is orthogonal.

## 5.4 Jordan Forms

In this section we are concerned with an $n \times n$ matrices that do not posess $n$ linearly independent eigenvectors. This was the situation in Example 2 above. If $\mathbf{A}$ does not possess $n$ linearly independent eigenvectors then there exists a non-singular matirx $\mathbf{P}$ such that,

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}$$

where $\mathbf{J}$ the Jordan matrix has the following structure,

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & 0 & \dots & 0 \\ 0 & \mathbf{J}_2 & 0 \dots & 0 \\ 0 & 0 & \mathbf{J}_3 \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \mathbf{J}_n \end{pmatrix}$$

where $\mathbf{J}_i$ are the Jordan blocks which have $\lambda$'s on the diagonal and $1$'s on the first superdiagonal. A typical form for a $(3 \times 3)$ Jordan block is illustrated below,

$$\mathbf{J}_i = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

$\mathbf{J}$ is also referred to as the Jordan canonical form.

If an $n \times n$ matrix $\mathbf{A}$ has $k$ linearly independent eigenvectors then the matrix $\mathbf{P}$ is constructed using these $k$ eigenvectors as well as the remainining $n-k$ generalized eigenvectors. Before we embark on determining generalized eigenvectors we spend some time on the structure of the Jordan forms themselves.

### 5.4.1 Structure of the Jordan Block

The structure of the Jordan block is best illustrated using some examples. Consider a $(3 \times 3)$ matrix with multiplicity $m = 3$. We can then have three different situations depending on the number of linearly independent eigenvectors.

**Case 1**: $1$ eigenvector and 2 generalized eigenvectors and the Jordan matrix has 1 Jordan block.

$$\mathbf{J} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

**Case 2**: 2 eigenvectors and 1 generalized eigenvectors and the Jordan matrix has 2 Jordan blocks. The two forms of the Jordan matrix are,

$$\mathbf{J} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \quad \text{or} \quad \mathbf{J} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

**Case 3**: 3 eigenvectors and 0 generalized eigenvectors. This case reduces to the diagonal form $\Lambda$ and can be interpreted as having 3 Jordan blocks.

$$\mathbf{J} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

In the above illustrations, the Jordan block is identified as a partitioned matrix. We can make the following statement relating the number of Jordan blocks to the number of linearly independent eigenvectors. The number of Jordan blocks in the Jordan canonical form $\mathbf{J}$ correspond to the number of linearly independent eigenvectors of the matrix $\mathbf{A}$. Further from the the examples above, the number of 1's on the super-diagonal is equivalent to the number of generalized eigenvectors used to construct the matrix, $\mathbf{P}$.

### 5.4.2 Generalized Eigenvectors

In this section we illustrate the procedure for finding generalized eigenvectors for a matrix with deficient eigenvectors. Consider for example a $(3 \times 3)$ matrix with eigenvalue $\lambda$ having multiplicity 3 and 1 eigenvector $\mathbf{x}$. In this case we would like to obtain 2 generalized eigenvectors, $\mathbf{q}_1$ and $\mathbf{q}_2$. The situation corresponds to Case 1, above with the Jordan matrix having two 1's on the off-diagonal.

$$\begin{aligned} \mathbf{P} &= [\mathbf{x}, \mathbf{q}_1, \mathbf{q}_2] \\ \mathbf{AP} &= [\lambda\mathbf{x}, \mathbf{A}\mathbf{q}_1, \mathbf{A}\mathbf{q}_2] \\ \mathbf{PJ} &= [\mathbf{x}, \mathbf{q}_1, \mathbf{q}_2] \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \\ &= [\lambda\mathbf{x}, \mathbf{x} + \lambda\mathbf{q}_1, \mathbf{q}_1 + \lambda\mathbf{q}_2] \end{aligned}$$

Equating $\mathbf{AP} = \mathbf{PJ}$, in the above equations we obtain the following equations for the generalized eigenvectors $\mathbf{q}_1$ and $\mathbf{q}_2$,

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{q}_1 = \mathbf{x} \qquad \text{and} \qquad (\mathbf{A} - \lambda\mathbf{I})\mathbf{q}_2 = \mathbf{q}_1$$

We make the following observations. Unlike eigenvectors which are obtained as solutions to a homogeneous problem, generalized eigenvectors are obtained as solutions to inhomogeneous equations as given above. We consider the generalized eigenvector corresponding to situation in Case 2 given above. In this case the matrix $\mathbf{P}$ is constructed by using only one generalized eigenvector. Using the same procedure as above,

$$\begin{aligned}
\mathbf{P} &= [\mathbf{x}_1, \mathbf{x}_2, \mathbf{q}_1] \\
\mathbf{AP} &= [\lambda\mathbf{x}_1, \mathbf{A}\mathbf{x}_2, \mathbf{A}\mathbf{q}_1] \\
\mathbf{PJ} &= [\mathbf{x}_1, \mathbf{x}_2, \mathbf{q}_1] \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \\
&= [\lambda\mathbf{x}_1, \lambda\mathbf{x}_2, \mathbf{x}_2 + \lambda\mathbf{q}_1]
\end{aligned}$$

Equating $\mathbf{AP} = \mathbf{PJ}$, in the above equations we obtain the following equations for the generalized eigenvector $\mathbf{q}_1$,

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{q}_1 = \mathbf{x}_2$$

In the above example there are a number of different variations to obtain the generalized eigenvector. $\mathbf{P}$ can be constructed by interchanging the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. In this case the equation for the generalized eigenvector reduces to,

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{q}_1 = \mathbf{x}_1$$

Additionaly the $\mathbf{P}$ matrix can be constructed by taking linear combinations of $\mathbf{x}_1$ and $\mathbf{x}_2$. If $\mathbf{u} = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2$, then

$$\mathbf{P} = [\mathbf{x}_1, \mathbf{u}, \mathbf{q}_1]$$

Working through the same procedure as outlined above, the equation for the generalized eigenvector is

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{q}_1 = \mathbf{u} = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2.$$

Clearly the above procedure illustrates that the generalized eigenvector can be constructed in many ways and is hence non-unique. Regardless of the manner in which the generalized eigenvector is obtained, the structure of the Jordan matrix is unaltered. Since the generalized eigenvector must be obtained by solving an inhomogeneous equation the issue of solvability must be confronted. The last example illustrates the number of ways in which the right hand side vector can be chosen to meet the solvability criterion or equivalently arrive at an system of equations that yields a solution.

*Question: In the last example where the matrix has two eigenvectors, derive the equations for the generalized eigenvector assuming*

$$\mathbf{J} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

## 5.5 Initial Value Problems

Consider the linear IVP of the following form

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{b}(t) \tag{5.25}$$

with initial condition, $\mathbf{x}(t = 0) = \mathbf{x}_0$. In Eq 5.25, each element of $\mathbf{x}$ and $\mathbf{b}$ are functions of time and the matrix $\mathbf{A}$ consists of constant coefficients. Hence

$$\mathbf{x} = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1(t) \\ b_2(t) \\ \vdots \\ b_n(t) \end{pmatrix}$$

Such systems of IVPs occur in staged processes, batch reactors, process control and vibration analysis. We will solve the above equation using the similarity transform technique introduced in the previous section.

We first consider the situation where, $\mathbf{A}$ has $n$ linearly independent eigenvectors. In this case $\mathbf{A}$ is diagonalizable. Premultiplying Eq. 5.25 by $\mathbf{P}^{-1}$

$$\begin{aligned}\frac{d(\mathbf{P}^{-1}\mathbf{x})}{dt} &= \mathbf{P}^{-1}\mathbf{A}\mathbf{x} + \mathbf{P}^{-1}\mathbf{b}(t)\\ \frac{d(\mathbf{P}^{-1}\mathbf{x})}{dt} &= \mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{x} + \mathbf{P}^{-1}\mathbf{b}(t)\end{aligned}$$

In the last line above we have inserted $\mathbf{P}\mathbf{P}^{-1}$ after $\mathbf{A}$. If $\mathbf{P}^{-1}\mathbf{x} = \mathbf{y}$ and $\mathbf{P}^{-1}\mathbf{b}(t) = \mathbf{g}(t)$ then the above system can be rewritten as,

$$\frac{d\mathbf{y}}{dt} = \Lambda\mathbf{y} + \mathbf{g}(t) \tag{5.26}$$

with the following IC, $\mathbf{y}(t=0) \equiv \mathbf{y}_0 = \mathbf{P}^{-1}\mathbf{x}_0$. Using the integrating factor $e^{-\Lambda t}$, Eq 5.26 can be rewritten as,

$$\frac{d}{dt}(e^{-\Lambda t}\mathbf{y}) = e^{-\Lambda t}\mathbf{g}(t) \tag{5.27}$$

whose general solution is

$$e^{-\Lambda t}\mathbf{y}(t) = \int_0^t e^{-\Lambda\tau}\mathbf{g}(\tau)\,d\tau + \mathbf{c} \tag{5.28}$$

Using the initial condition $\mathbf{y}_0 = \mathbf{P}^{-1}\mathbf{x}_0$, Eq. 5.28 reduces to,

$$\mathbf{y}(t) = \int_0^t e^{\Lambda(t-\tau)}\mathbf{g}(\tau)\,d\tau + e^{\Lambda t}\mathbf{y_0} \tag{5.29}$$

The solution can be expressed in a more compact form as

$$\mathbf{x} = \mathbf{P}\mathbf{y} = \mathbf{P}[e^{\Lambda t}\mathbf{y}_0 + \mathbf{f}(t)] \quad \text{where} \quad \mathbf{f}(t) = \begin{pmatrix} \int_0^t e^{\lambda_1(t-\tau)}\,g_1(\tau)\,d\tau \\ \int_0^t e^{\lambda_2(t-\tau)}\,g_2(\tau)\,d\tau \\ \vdots \\ \int_0^t e^{\lambda_n(t-\tau)}\,g_n(\tau)\,d\tau \end{pmatrix}$$

In order to obtain the solution we need to obtain an expression for $e^{\Lambda t}$. This can be obtained in the following manner. Expanding $e^{\Lambda t}$ in a Taylor series,

$$e^{\Lambda t} = \mathbf{I} + \Lambda t + \frac{(\Lambda t)^2}{2!} + \frac{(\Lambda t)^3}{3!} + \ldots,$$

Substituting $\Lambda$ in the above expression and collecting terms,

$$e^{\Lambda t} = \begin{pmatrix} \sum_{n=0}^{\infty}\frac{(\lambda_1 t)^n}{n!} & 0 & \ldots & 0 \\ 0 & \sum_{n=0}^{\infty}\frac{(\lambda_2 t)^n}{n!} & 0 & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \sum_{n=0}^{\infty}\frac{(\lambda_n t)^n}{n!} \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2 t} & 0 & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & e^{\lambda_n t} \end{pmatrix}$$

If $\mathbf{A}$ is not diagonalizable then the above solution procedure remains unaltered. However $\Lambda$ is replaced with $\mathbf{J}$. In this case we need to obtain an expression for $e^{\mathbf{J}t}$. Consider the following example. If

$$\mathbf{J} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \tag{5.30}$$

then $\mathbf{J} = \Lambda + \mathbf{S}$, where $\Lambda$ is the diagonal matrix and $\mathbf{S}$ is a matrix containing the off-diagonal terms. Then

$$e^{\mathbf{J}t} = e^{\Lambda t} e^{\mathbf{S}t} \quad \text{where} \quad \Lambda = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

where $e^{\Lambda t}$ is evaluated as illustrated above. In order to evaluate $e^{\mathbf{S}t}$ we proceed in a similar manner and carry out a Taylor expansion,

$$e^{\mathbf{S}t} = \mathbf{I} + \mathbf{S}t + \frac{(\mathbf{S}t)^2}{2!} + \ldots, \tag{5.31}$$

Due to the structure of matrix $\mathbf{S}$, the number of terms that are retained in the Taylor expansion is only 3 as the powers of $\mathbf{S}$ greater than 2 are identically zero. The reader should check this. Collecting terms in the expansion given in Eq. 5.31 we obtain

$$e^{\mathbf{S}t} = \begin{pmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad e^{\mathbf{J}t} = \begin{pmatrix} e^{\lambda t} & te^{\lambda t} & \frac{t^2}{2}e^{\lambda} t \\ 0 & e^{\lambda t} & te^{\lambda t} \\ 0 & 0 & e^{\lambda t} \end{pmatrix}$$

The above procedure can be generalized for an $(n \times n)$ Jordan matrix of the form given in Eq. 5.30 and,

$$e^{\mathbf{J}t} = \begin{pmatrix} e^{\lambda t} & te^{\lambda t} & \ldots, & \frac{t^{n-1}}{(n-1)!}e^{\lambda} t \\ 0 & e^{\lambda t} & \ldots, & \frac{t^{n-2}}{(n-2)!}e^{\lambda} t \\ 0 & 0 & \ldots, & \vdots \\ 0 & 0 & & e^{\lambda t} \end{pmatrix}.$$

## 5.6 Eigenvalues and Solutions of Linear Equations

While solving linear equations, $\mathbf{Ax} = \mathbf{b}$ it is important to understand the sensitivity of the solution to small changes in the coefficients of the matrix $\mathbf{A}$ or the elements in the vector $\mathbf{b}$.

The sensitivity usually arises from round-off error during a numerical solution such as Gauss elimination. A measure of the sensitivity to small perturbations is known as the condition number of the matrix. Hence a matrix whose solutions are sensitive to small changes in the coefficients is said to be poorly conditioned. We will use concepts of matrix and vector norms to quantify these concepts and connect this issue of sensitivity to the eigenvalues of the matrix.

**Normed Space** The norm is simply the notion of length that we have encountered while discussing inner product spaces. More formally, $\|\mathbf{x}\|$ is said to be a norm on a linear space $X$, $\mathbf{x}, \mathbf{y} \in X$ if it satisfies the following properties,

$$\begin{aligned}
&(i) && \|\mathbf{x}\| > 0 \\
&(ii) && \|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \qquad \text{Triangular Inequality} \\
&(iii) && \|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\| \\
&(iv) && \|\mathbf{x}\| = 0 \qquad \text{If and only if } \mathbf{x} = 0
\end{aligned}$$

Some examples of commonly encountered norms are

The 2 norm

$$\|\mathbf{x}\|_2 = \left[\sum_{i=1}^{n} |x_i|^2\right]^{1/2}$$

The p norm

$$\|\mathbf{x}\|_p = \left[\sum_{i=1}^{n} |x_i|^2\right]^{1/p}, \quad 1 \leq p < \infty$$

The $\infty$ norm

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

The norm incorporates the definition of a distance function or metric, $d(x, y)$

$$d(x, y) = \|x - y\|$$

If $x = (x_1, x_2)$ and $y = (y_1, y_2)$, then

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

which is the familiar example of the distance in $\mathcal{R}^2$.

**Matrix Norms**: $\|\mathbf{A}\|$ is a matrix norm for a matrix $\mathbf{A}$ if it satisfies the following properties of a normed space,

$$
\begin{aligned}
&(i) \quad \|\mathbf{A}\| > 0 \\
&(ii) \quad \|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\| \qquad \text{Triangular Inequality} \\
&(iii) \quad \|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\| \\
&(iv) \quad \|\mathbf{A}\| = 0 \qquad \text{If and only if } \mathbf{A} = 0
\end{aligned}
$$

Further

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$$

The matrix norm is compatible with a vector norm if

$$\|\mathbf{Ax}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|$$

Examples of commonly encountered matrix norms are given below,

The 1 norm, or the maximum column sum,

$$\|\mathbf{A}\|_1 = \max_{1\leq j\leq n} \left[\sum_{i=1}^{n} |a_{ij}|\right]^{1/2}$$

The $\infty$ norm or the maximum row sum,

$$\|\mathbf{A}\|_\infty = \max_{1\leq i\leq n} \left[\sum_{j=1}^{n} |a_{ij}|\right]^{1/2}$$

The spectral norm,

$$\|\mathbf{A}\|_2 = [\rho(\mathbf{A}^*\mathbf{A})]^{1/2}$$

where $\rho(\mathbf{A})$ is the spectral radius of $\mathbf{A}$ defined as the maximum eigenvalue of $\mathbf{A}$. If $\mathbf{A}$ is Hermitian, $\mathbf{A}^* = \mathbf{A}$ and

$$\|\mathbf{A}\|_2 = |\lambda_{max}|$$

If $\mathbf{Ax} = \lambda\mathbf{x}$ then any norm of $\mathbf{A}$ is an upper bound on the eigenvalues.

$$|\lambda|\|\mathbf{x}\| = \|\lambda\mathbf{x}\| = \|\mathbf{Ax}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|$$

and

$$|\lambda| \leq \|\mathbf{A}\|$$

**Errors and Perturbation**

Consider the linear equation,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{5.32}$$

If $\delta\mathbf{b}$ is a small pertubation to the vector $\mathbf{b}$ then let $\delta\mathbf{x}$ be the corresponding pertubation to the solution vector $\mathbf{x}$ and

$$\mathbf{A}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b} \tag{5.33}$$

The problem lies in determing a bound on the perturbation to the solution vector $\mathbf{x}$. Expanding, Eq. 5.33 and noting that $\mathbf{A}\mathbf{x} = \mathbf{b}$,

$$\mathbf{A}\delta\mathbf{x} = \delta\mathbf{b} \tag{5.34}$$

and

$$\delta\mathbf{x} = \mathbf{A}^{-1}\delta\mathbf{b} \tag{5.35}$$

From Eq. 5.34

$$\|\delta\mathbf{b}\| = \|\mathbf{A}\delta\mathbf{x}\| \leq \|\mathbf{A}\|\|\delta\mathbf{x}\| \tag{5.36}$$

and from Eq. 5.35,

$$\|\delta\mathbf{x}\| = \|\mathbf{A}^{-1}\delta\mathbf{b}\| \leq \|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\| \tag{5.37}$$

From $\mathbf{A}\mathbf{x} = \mathbf{b}$

$$\|\mathbf{b}\| = \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\|\|\mathbf{x}\| \tag{5.38}$$

Combining Eqs. 5.37 and 5.38

$$\|\delta\mathbf{x}\|\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\|\|\mathbf{x}\| \tag{5.39}$$

If $\|\mathbf{b}\| \neq 0$ then Eq. 5.39 reduces to,

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{5.40}$$

The condition number $\kappa(\mathbf{A})$ is defined as

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| \tag{5.41}$$

and Eq. 5.42 is,

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa(\mathbf{A}) \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} \tag{5.42}$$

If the 2 or spectral norm is used and the matrix is symmetric, then

$$\kappa(\mathbf{A}) = \frac{|\lambda_{max}|}{|\lambda_{min}|} \tag{5.43}$$

Eq. 5.42 indicates that if the condition number is large then small perturbations in the vector $\mathbf{b}$ can amplify the errors in the solution vector $\mathbf{x}$. Matrices that are nearly singular (i.e with one eigenvalue close to zero) are clearly poorly conditioned. There are many numerical methods developed to improve the conditioning of matrices. Clearly precision related conditioning can be alleviated to some extent by using higher precision computing. While deriving the bounds as given in Eq. 5.42 we assumed that the errors occurred only in the vector $\mathbf{b}$. We next consider, the situation where the error occurs in the matrix.

### 5.6.1 Positive Definite Matrices

A matrix is $\mathbf{A}$ is said to be positive definite if

$$< \mathbf{A}\mathbf{x}, \mathbf{x} > \quad > 0 \quad \text{for} \quad \mathbf{x} \neq 0$$

If $\mathbf{A}$ is symmetric then $\mathbf{A}$ is said to be *symmetric positive definite* (SPD). To show that the eigenvalues of a positive definite matrix are always positive;

$$< \mathbf{A}\mathbf{x}, \mathbf{x} >= \lambda < \mathbf{x}, \mathbf{x} >= \lambda \|\mathbf{x}\|^2 > 0$$

Hence all $\lambda$'s are positive. As a consequence, the determinant of a positive definite matrix is non-zero. If $\mathbf{A}$ is singular, then $\exists$ a nonzero vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x}$ = 0, which implies that $< \mathbf{A}\mathbf{x}, \mathbf{x} >= 0$.

**Spectral Radius** The spectral radius, $\rho(\mathbf{A})$ of a matrix $\mathbf{A}$ is the maximum value of the modulus of its eigenvalues.

$$\rho(\mathbf{A}) = \max_i \ |\lambda_i|$$

There are several localization theorems which yield information on the bounds for the eigenvalues. The most important theorem is the Gerschgorin's theorem.

**Gerschgorin's Theorem:** Let $\mathbf{A}$ be a general $n \times n$ matrix whose eigenvalues can be either real or complex. If

$$r_i = \sum_{j=1, j\neq i}^{n} |a_{ij}| \quad i = 1 \ldots, n$$

is the sum of the off-diagonal elements in the $i^{th}$ row. Let $\mathcal{D}_i$ be the disk in the complex plane of radius $r_i$ and centered at $a_{ii}$,

$$\mathcal{D}_i = \{z : |z - a_{ii}| \leq r_i\} \quad i = 1, \ldots, n$$

Gerschogorin's theorem, states that all eigenvalues of $\mathbf{A}$ lie in the union of disks $\mathcal{D}_i$. Thus

$$\lambda_i \in \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3 \ldots \cup \mathcal{D}_n \quad i = 1, \ldots, n$$

**Proof** Consider any $\lambda$ with corresponding eigenvector $\mathbf{x}$. The eigen equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ can be expressed as,

$$(\lambda - a_{ii})\mathbf{x}_i = \sum_{j=1, j\neq i}^{n} a_{ij}x_j \quad \text{for} \quad i = 1, \ldots n \tag{5.44}$$

where $x_j$ is the j$^{th}$ component in the eigenvector $\mathbf{x}$. Let $x_k$ be the component with the largest absolute value in the vector $\mathbf{x}$. Then $|x_j|/|x_k| \leq 1$ for $j = 1, \ldots n$. Eq. 5.44 for $i = k$ can be expressed as,

$$(\lambda - a_{kk}) = \sum_{j=1, j\neq k}^{n} a_{kj}\frac{x_j}{x_k} \tag{5.45}$$

Taking moduli on both sides,

$$|\lambda - a_{kk}| \leq \sum_{j=1, j\neq k}^{n} |a_{kj}|\frac{|x_j|}{|x_k|} \leq \sum_{j=1, j\neq k}^{n} |a_{kj}| = r_k$$

Thus $\lambda$ is contained in the disk $\mathcal{D}_k$ centered at $a_{kk}$. A similar procedure follows for all the $\lambda$'s. Hence the eigenvalues lie in the union of the disks, $\mathcal{D}_k$ $k = 1, \ldots n$

### 5.6.2 Convergence of Iterative Methods

The spectral radius of a matrix is useful while analyzing convergence of iterative processes. We will show that sequence of vectors generated by the iterative process,

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k \qquad k = 0, \ldots \tag{5.46}$$

will tend to zero if and only if, $\rho(\mathbf{A}) < 1$. $\mathbf{x}_0$ is an arbitrary initial vector. Further if $\rho(\mathbf{A}) > 1$ then the sequence will diverge. The convergence of the limiting vector can be analyzed by expanding in a basis made up of eigenvectors, $\{\mathbf{u}_i\}$ of $\mathbf{A}$.

$$\begin{aligned}
\mathbf{x}_0 &= \sum_{i=1}^{n} \alpha_i \, \mathbf{u}_i \\
\mathbf{x}_1 &= \mathbf{A}\mathbf{x}_0 = \sum_{i=1}^{n} \alpha_i \, \mathbf{A}\mathbf{u}_i = \sum_{i=1}^{n} \alpha_i \, \lambda_i \mathbf{u}_i \\
\mathbf{x}_2 &= \mathbf{A}\mathbf{x}_1 = \sum_{i=1}^{n} \alpha_i \, \lambda_i \mathbf{A}\mathbf{u}_i = \sum_{i=1}^{n} \alpha_i \, \lambda_i^2 \mathbf{u}_i \\
&\vdots \qquad \vdots \\
\mathbf{x}_k &= \sum_{i=1}^{n} \alpha_i \, \lambda_i^{k+1} \mathbf{u}_i
\end{aligned}$$

Since $\rho(\mathbf{A}) < 1$, the powers of $\lambda_i$ tend to zero as $k \to \infty$ and

$$\lim_{k\to\infty} \mathbf{x}_k = 0$$

Alternately the iterative process can be analyzed by examining the powers of the matrix $\mathbf{A}$. Thus,

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{A}\mathbf{x}_0 \\
\mathbf{x}_2 &= \mathbf{A}\mathbf{x}_1 = \mathbf{A}^2\mathbf{x}_0 \\
&\vdots \qquad \vdots \\
\mathbf{x}_k &= \mathbf{A}^k\mathbf{x}_0
\end{aligned}$$

If $\mathbf{A}$ can be diagonalized using similarity transforms, then

$$\mathbf{A}^k = \mathbf{P}^{-1}\mathbf{\Lambda}^k\mathbf{P}$$

where $\Lambda$ is the diagonal matrix with eigenvalues on the diagonal, and

$$\lim_{k\to\infty} \mathbf{A}^k = 0 \quad \text{since} \quad \lim_{k\to\infty} \mathbf{\Lambda}^k = 0 \quad \text{if} \quad \rho(\mathbf{A}) < 1$$

and

$$\lim_{k\to\infty} \mathbf{x}_k = 0$$

We leave it as an exersize to show that the limiting vector $\mathbf{x}_k$ will converge to zero when,

$$\mathbf{A}^k = \mathbf{P}^{-1}\mathbf{J}^k\mathbf{P}$$

where $\mathbf{J}$ is the Jordan canonical form, i.e.,

$$\lim_{k\to\infty} \mathbf{J}^k = 0 \quad \text{if} \quad \rho(\mathbf{A}) < 1$$

We illustrate the application of these ideas with the analysis of convergence properties of iterative methods such as the Jacobi and Gauss-Seidel methods used for solutions of linear equations which results in large sparse matrices. We briefly outline the procedure for the solution to $\mathbf{Ax} = \mathbf{b}$ using these methods.

**Jacobi's Method** Consider the solution to $\mathbf{Ax} = \mathbf{b}$ using the Jacobi's method. Rewrite

$$\mathbf{A} = \mathbf{D} - \mathbf{B} \tag{5.47}$$

where

$$\mathbf{D} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \quad \mathbf{B} = -\begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ a_{21} & 0 & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Substituting, $\mathbf{A} = \mathbf{D} - \mathbf{B}$ into $\mathbf{Ax} = \mathbf{b}$,

$$\mathbf{Dx} = \mathbf{Bx} + \mathbf{b}$$

This can be solved iteratively using the following numerical scheme,

$$\mathbf{x}_{k+1} = \mathbf{D}^{-1}(\mathbf{Bx}_k + \mathbf{b}) \quad k = 1, \dots \tag{5.48}$$

and the solution is the limiting vector as $k \to \infty$.

**The Gauss Seidel Method** This is an improvement over the Jacobi's method as it uses the latest updated components of the vector during each iteration. Here

$$\mathbf{A} = -(\mathbf{L} + \mathbf{U}) + \mathbf{D} \tag{5.49}$$

where $\mathbf{D}$ is the diagonal matrix as defined above and $\mathbf{L}$ is a lower triangular matrix and $\mathbf{U}$ is the upper triangular matrix as shown below,

$$\mathbf{L} = -\begin{pmatrix} 0 & 0 & \dots & 0 \\ a_{21} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,n-1} & 0 \end{pmatrix} \quad \mathbf{U} = -\begin{pmatrix} 0 & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{n-1,n} \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Substituting, Eq. 5.49 into $\mathbf{Ax} = \mathbf{b}$ and rearranging,

$$\mathbf{x}_{k+1} = (\mathbf{D} - \mathbf{L})^{-1}(\mathbf{U}\mathbf{x}_k + \mathbf{b}) \quad k = 1, \ldots$$

Convergence of these iterative methods can be analyzed in the following manner. We illustrate the analysis with the Jacobi's method. Let $\mathbf{x}_0$ be the exact solution to $\mathbf{Ax} = \mathbf{b}$. Then, Eq 5.48 is

$$\mathbf{x}_0 = \mathbf{D}^{-1}(\mathbf{B}\mathbf{x}_0 + \mathbf{b}) \tag{5.50}$$

Note that $\mathbf{x}_0$ is also referred to as a fixed point of the mapping. Fixed points will be discussed in Chapter **??**. Subtracting Eq. 5.48 from 5.50,

$$\mathbf{b}_{k+1} = \mathbf{H}\mathbf{b}_k \tag{5.51}$$

where $\mathbf{b}_{k+1}$ is the error vector, $\mathbf{x}_0 - \mathbf{x}_{k+1}$ at iterate $k+1$ and $\mathbf{H} = \mathbf{D}^{-1}\mathbf{B}$. Eq. 5.51 is of the same form as Eq. 5.46 and $\mathbf{b}$ will tend to zero as $k \to \infty$ when the spectral radius, $\rho(\mathbf{H}) < 1$. The conditions for the $\rho(\mathbf{H}) < 1$ can be obtained by examining the spectral radii of the Gerschgorin discs. The matrix,

$$\mathbf{H} = \mathbf{D}^{-1}\mathbf{B} = -\begin{pmatrix} 0 & a_{12}/a_{11} & a_{13}/a_{11} & \ldots & a_{1n}/a_{11} \\ a_{21}/a_{22} & 0 & a_{23}/a_{22} & \ldots & a_{2n}/a_{22} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1}/a_{nn} & a_{n2}/a_{nn} & \ldots & a_{n-1,1}/a_{nn} & 0 \end{pmatrix} \tag{5.52}$$

Since the radii, $r_i$ of the Gerschgorin discs are the sum of the off-diagonal elements of $\mathbf{H}$,

$$r_i = \sum_{j=1, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|} \quad i = 1 \ldots n \tag{5.53}$$

For $\rho(\mathbf{H}) < 1$,

$$r_i = \sum_{j=1, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|} < 1 \quad i = 1 \ldots n \tag{5.54}$$

which yields the following condition,

$$|a_{ii}| > \sum_{j=1, j\neq i}^{n} |a_{ij}| \tag{5.55}$$

Matrices which satisfy the condition given by Eq. 5.55 are referred to as *strictly diagonally dominant matrices* which will result in converged solutions to the Jacobi method, regardless of

the initial vector $\mathbf{x}_0$ used in the iteration. Clearly a good guess of the starting vector will reduce the number of iterations required to obtain the solution. Note that the result only comments on whether the iteration will converge or not. This is an instructive illustration of the utility of the Gerschgorin's theorem. The convergence criterion for the Gauss-Seidel method is left as an exercize.

## 5.7 Summary

The main goal of this Chapter was to analyze the eigenvalue-eigenvector problem for a matrix. Once the eigenvalues are obtained, the main problem reduces to finding the eigenvectors and understanding the properties between the eigenvectors themselves. Theorem 2 provides the foundation for constructing orthogonal sets of eigenvectors for Hermitian matrices. The theorem was discussed with relevance to matrices, however the generality of the theorem for differential and intergral operators has far reaching consequences, laying the foundation for obtaining orthogonal eigenfunctions for differential operators and developing a theory of Fourier series. These connections will be drawn in later Chapters. We also discussed the significance of using the eigenvectors as a basis set and illustrated its utility in solving $\mathbf{Ax} = \mathbf{b}$. In the last part of this Chapter we discussed similarity transforms and its utility in working with functions of matrices as well as solutions of linear initial value problems. In this context we introduced the Jordan canonical form and described a new set of vectors called generalized eigenvectors. The reader should realize that geneneralized eigenvectors are required for non-symmetric matrices where only a partial set of eigenvectors can be found.

## PROBLEMS

1. **Skew-Symmetric Matrix**

   A matrix $\mathbf{A}$ is said to be skew symmetric or skew self-adjoint if $\mathbf{A} = -\mathbf{A}^*$. Show that the eigenvalues are imaginary (or zero) and that eigenvectors corresponding to distinct eigenvectors are orthogonal.

2. **Normal Matrices**

   If $\mathbf{A}\mathbf{A}^*$ = $\mathbf{A}^*\mathbf{A}$, then $\mathbf{A}$ is said to be normal.

   (a) Show that for any complex number $\alpha$,

   $$||\mathbf{A}\mathbf{x} - \alpha\mathbf{x}|| = ||\mathbf{A}^*\mathbf{x} - \alpha^*\mathbf{x}||$$

   (b) If $\mathbf{z}$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda$ show that it is also an eigenvector of $\mathbf{A}^*$. What is the corresponding eigenvalue of $\mathbf{A}^*$ ?

   (c) Let $\lambda = \mu + i\nu$ be an eigenvalue of $\mathbf{A}$ with eigenvector $\mathbf{z}$. First show that $\mathbf{A}$ can be decomposed in the following manner,

   $$\mathbf{A} = \mathbf{A}_R + i\mathbf{A}_I,$$

   where $\mathbf{A}_R = \mathbf{A^*}_R$ and $\mathbf{A}_I = \mathbf{A^*}_I$. Next show that $\mathbf{z}$ is an eigenvector of $\mathbf{A}_R$ and $\mathbf{A}_I$ with eigenvalues $\mu$ and $\nu$ respectively.

3. Consider a $(4 \times 4)$ matrix with one multiple eigenvalue. Write out the possible Jordan canonical forms.

4. **Symmetric Matrix**

   Consider the following matrix

   $$\mathbf{A} = \begin{pmatrix} 7 & -16 & -8 \\ -16 & 7 & 8 \\ -8 & 8 & -5 \end{pmatrix}$$

   (a) Find the eigenvalues and eigenvectors of $\mathbf{A}$?

   (b) Find a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ where $\mathbf{b} = \{1, 2, 1\}$ by expanding $\mathbf{x}$ in the normalized eigenvectors of $\mathbf{A}$.

5. Consider the following matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}$$

   (a) Find the eigenvalues and eigenvectors of A?

   (b) Find a solution to $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{b} = \{0, -2, 3\}$ by expanding $\mathbf{x}$ in the eigenvectors of $\mathbf{A}$.

6. **Solvability Conditions**

   Consider the non-homogeneous equation

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{b}, \tag{5.56}$$

   where **A** is a square matrix of dimension $n$. Let

$$\mathbf{u} = \sum_{j=1}^{n} c_j \phi_j, \tag{5.57}$$

   where $c_j$'s are the coefficients of the expansion and $\phi_j$'s are the eigenvectors of $\mathbf{A}$.

   (a) If $\mathbf{A}$ is self adjoint and $\lambda$ in Eq. 5.56 is not an eigenvalue of $\mathbf{A}$, then obtain an expression for the coefficients $c_j$ in the expansion. What are the solvability conditions for Eq. 5.56 (Fredholms Alternative Theorem)?

   (b) Re-work part a) for the case when $\lambda$ is a particular eigenvalue of $\mathbf{A}$. Note how the solvability conditions are connected to the eigenfunctions of $\mathbf{A}$.

   (c) If $\mathbf{A}$ is a non self adjoint matrix with n linearly independent eigenvectors then the solution results in

$$\mathbf{Mc} = \mathbf{f}.$$

   Write out the components of the matrix $\mathbf{M}$ and vector $\mathbf{f}$

7. **IVP**

   Using similarity transforms solve the system

$$\frac{d\mathbf{x}}{dt} = \mathbf{Ax}$$

with initial conditions $\mathbf{x}(t=0) = \{1, 1, 1\}$ where

$$\mathbf{A} = \begin{pmatrix} 5 & -3 & -2 \\ 8 & -5 & -4 \\ -4 & 3 & 3 \end{pmatrix}$$

8. Let

$$\mathbf{A} = \begin{pmatrix} -2 & 1 \\ -1 & -2 \end{pmatrix}$$

   (a) Find the eigenvalues and eigenvectors of $\mathbf{A}$.

   (b) Do the eigenvectors form an orthogonal set ?

   (c) Using similarity transforms obtain the solution to

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u} + \mathbf{b}$$

   where $\mathbf{b} = \{1, 1\}$ and $\mathbf{u}(t=0) = \{0, 0\}$.

   (d) How does your solution behave as $t$ tends to $\infty$ ?

9. **Skew Symmetric System**

   Using similarity transforms solve the system

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{b}(\mathbf{t})$$

   with initial conditions $\mathbf{x}(t=0) = \{1/\sqrt{2}, 1/\sqrt{2}, 1\}$ where

$$\mathbf{A} = \begin{pmatrix} -i & i & 0 \\ i & -i & 0 \\ 0 & 0 & -i \end{pmatrix}, \ \mathbf{b}(\mathbf{t}) = \begin{pmatrix} \sqrt{2}t \\ \sqrt{2}t \\ \exp(-t) \end{pmatrix}$$

   Comment on the asymptotic stability of the system.

10. Consider the initial value problem

$$\frac{d^2u}{dt^2} + 5\frac{du}{dt} + 6u = e^{-t} \tag{5.58}$$

    with initial condition, $u(t=0) = u'(t=0) = 1$

(a) Reduce the above ode to a set of first order linear differential equations and represent them in matrix vector form,

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u} + \mathbf{b}(t) \tag{5.59}$$

Write out the components for the matrix $\mathbf{A}$ and vectors $\mathbf{u}$, $\mathbf{b}(t)$ and initial condition $\mathbf{u}(t=0)$. Obtain the solution to Eq. 5.59 using similarity transformations.

(b) The above solution can also be solved using the corresponding Greens function, $g(t,\xi)$, for the second order differential operator given in Eq. 5.58. Using the Green's functions the solution to Eq. 5.58 can be expressed as

$$u(t) = c_1 u_1(t) + c_2 u_2(t) + \int_o^t g(t,\xi) e^{-\xi} d\xi \tag{5.60}$$

where the Green's function $g(t,\xi) = exp[2(\xi - t)] - exp[3(\xi - t)]$. $u_1(t)$ and $u_2(t)$ are two linearly independent solutions to the homogeneous differential equation,

$$\frac{d^2u}{dt^2} + 5\frac{du}{dt} + 6u = 0 \tag{5.61}$$

Using Eq. 5.60 find the solution $u(t)$ for initial condition, $u(t=0) = u'(t=0) = 1$, i.e find $u_1(t)$, $u_2(t)$ and the constants $c_1$ and $c_2$. You will have to evaluate the integral in Eq. 5.60 to obtain the complete solution.

11. Consider the following non-symmetric matrix with real coefficients,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

(a) Derive the conditions on the coefficients $a_{ij}$, when the matrix has two similar eigenvalues.

(b) In this situation show that the matrix can possess only one eigenvector. Derive a general expression for the eigenvector, $\mathbf{x}$ in terms of $a_{ij}$

(c) Write out the equation that will be used to obtain the generalized eigenvector, $\mathbf{q}$. Write out the form for the Jordan matrix.

(d) Show that when the matrix has two similar eigenvalues the generalized eigenvector can always be obtained (i.e the solvability criterion are always satisfied).

(e) Using similarity transforms obtain a solution to the following initial value problem,

$$\begin{aligned} dx_1/dt &= 3x_1 + x_2 \\ dx_2/dt &= -x_1 + x_2 \end{aligned}$$

for the initial conditions $x_1 = 1, t = 0$ and $x_2 = 1, t = 0$. Is the system asymptotically stable. Why?

(f) Qualitatively sketch your solutions.

12. **IVPs**

Consider the Initial Value Problem;

$$\frac{d^3x}{dt^3} + a_1\frac{d^2x}{dt^2} + a_2\frac{dx}{dt} + a_3x = f(t)$$

with initial conditions $x(0) = x'(0) = x''(0) = 0$. Reduce this to a system of first order equations of the form

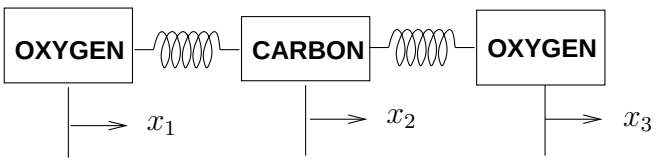
$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u} + \mathbf{b}$$

(a) If $a_3 = 0$, what are the conditions on $a_1$ and $a_2$ for the system to have a stable solution.

(b) If $a_1 = -4$, $a_2 = 3$ and $f(t) = sin(t)$ obtain a solution to the IVP using the similarity transform method.

13. **Normal Mode Analysis:Vibration of a $CO_2$ Molecule**

Consider a spring and mass model of a $CO_2$ molecule as shown in the figure below. The oxygen molecules have mass $m_o$ and the carbon molecule has mass $m_c$. The springs have a spring constant $k$ and obey Hooke's law. Using Newton's laws and assuming




that the motion is constrained along the $x - axis$ the system of equations describing the

displacement of masses is

$$\begin{aligned}\frac{d^2x_1}{dt^2} &= -a(x_1 - x_2)\\ \frac{d^2x_2}{dt^2} &= -b(x_2 - x_1) - b(x_2 - x_3)\\ \frac{d^2x_3}{dt^2} &= -a(x_3 - x_2)\end{aligned}$$

where $a = k/m_o$ and $b = k/m_c$

(a) Assuming a solution of the form

$$x_n(t) = x_n e^{i\sqrt{\omega}t} \qquad n = 1, 2, 3$$

where $i = \sqrt{-1}$ and $\omega$ is a natural frequency of oscillation of the system reduce the set of ode's to an eigenvalue problem of the form

$$\mathbf{A}\mathbf{x} = \omega\mathbf{x}.$$

(b) Find the eigenvalues $\omega$.

(c) Find the corresponding eigenvectors.

(d) Noting that the components of the eigenvectors correspond to the displacement of the molecules, give a physical explanation for eigenvectors.

14. **Projection Theorem** Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$$

where $i = \sqrt{-1}$.

(a) Find the eigenvalues and normalized eigenvectors of $\mathbf{A}$.

(b) Find the projections $\mathbf{P}_1$ and $\mathbf{P}_2$ of $\mathbf{A}$.

(c) Using the projection theorem evaluate $\mathbf{A}^2$ and $e^{\mathbf{A}t}$.

# Chapter 6

# Solutions of Non-Linear Equations

Non-linear differential and algebraic equations arise in a wide variety of engineering situations and we have seen some examples of non-linear operators in Chapter **??**. Numerical solutions of non-linear differential equations result in a set of non-linear algebraic equations. Although there are a number of techniques available for solving non-linear algebraic equations, in this Chapter we will focus on primarily two methods, the Picard and Newton-Raphson methods. The primary goal here is to develop a framework to analyze non-linear equations. A large number of excellent texts cover the variety of numerical methods available for solving non-linear equations. In order to formally treat non-linear equations and discuss their convergence, existence and uniqueness aspects, we need to introduce the metric space. In many situations we can express non-linear or linear equations in the following implicit manner,

$$u = Lu$$

where $L$ can either be a linear or non-linear operator and $u$ is the unknown we seek. Examples of equations that can be cast in the form of Eq. 6 are

1. $$x = \tan x$$

2. $$x = x^2 + \sin x + 2$$

3. $$u(x) = \int_0^x k(x, y)u(y)\, dy$$

4.

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

**Fixed Points:** $u$ is said to be a fixed point of the mapping $L$ if

$$u = Lu$$

Thus $L$ operating on $u$ leaves it unchanged and is a solution to $F(u) = u - Lu = 0$. The method of successive substitution can be used to determine the fixed point in the following manner,

$$u_{n+1} = Lu_n \quad n = 0 \ldots,$$

If $\overline{u}$ is a fixed point of $L$ then

$$\lim_{n\to\infty} u_n = \overline{u}$$

and $\overline{u} = L\overline{u}$. This method of successive substitution also known as Picard's iterative method will work only for a certain class of mappings or operators referred to as contractions. As the name implies the mapping contracts the distance between successive iterates and the generated sequence in Eq. 6 tends to a limiting value $v$ called the fixed point under certain conditions. Before introducing the contraction mapping theorem let us formally define the metric space which provides a framework for defining distances between elements in a space.

**Metric Space** $(X, d)$ is said to be a metric space if the distance between any two points $x$ and $y$ in $X$ denoted by $d(x, y)$ statisfies the following axioms More formally,

$(i)$ $d(x, y) \geq 0,\ d(x, y) = 0 \Rightarrow x = y$ **Positivity**

$(ii)$ $d(x, y) = d(y, x)$ Symmetry

$(ii)$ $d(x, y) \leq d(x, z) + d(z, y)\ x, y, z \in X$ Triangular Inequality

Thus the metric, $d(x, y)$ is simply a distance function and hence a scalar quantity. Some examples of commonly encountered metrics are given below

If $x$ and $y$ are two vectors in $\mathcal{R}^n$,

$$d(x, y) = \left[\sum_{i=1}^{n} (x_i - y_i)^2\right]^{1/2}$$

If $x = (x_1, x_2)$ and $y = (y_1, y_2)$, then

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

which is the familiar example of the distance in $\mathcal{R}^2$, called the Euclidean distance. This metric is used frequently in least squares fitting of data. The $p$ metric is a more general definition of the metric,

$$d_p(x, y) = \left[ \sum_{i=1}^{n} |x_i - y_i|^p \right]^{1/p}, \quad 1 \leq p < \infty$$

The $\infty$ metric

$$d_\infty(x, y) = \max_{1 \leq i \leq n} |x_i - y_i|$$

The $\infty$ metric is useful in many engineering situations. While determining the uniformity of temperature in an object the difference between the maximum and minimum temperatures is an example of $d_\infty$. The metric is related to the norm in the following manner,

$$d(x, y) = \|x - y\|$$

If $f(x)$ and $g(x)$ are two continuous functions in $C[a, b]$, then

$$d_p(f, g) = \left[ \int_a^b |f(x) - g(x)|^p \, dx \right]^{1/p}, \quad 1 \leq p < \infty$$

Example: To show that $(X, d)4$ is a valid metric space the distance function must satisfy the axioms of the metric space. We illustrate this with the metric defined above for finite sums,

$$d_p(x, y) = \left[ \sum_{i=1}^{n} |x_i - y_i|^p \right]^{1/p}, \quad 1 \leq p < \infty$$

It is easy to see that the postivity and symmetry properties of the metric are satisfied. In order to prove that $d_p(x, y)$ satisfies the triangular inequality we need to use the Minkowski Inequality forfinite sums[**?**],

$$\left\{ \sum_{i=1}^{n} |x_i \pm y_i|^p \right\}^{1/p} \leq \left\{ \sum_{i=1}^{n} |x_i|^p \right\}^{1/p} + \left\{ \sum_{i=1}^{n} |y_i|^p \right\}^{1/p}$$

Then,

$$\begin{aligned} d(x,y) &= \left\{\sum_{i=1}^{n}|x_i-y_i|^p\right\}^{1/p} \\ &= \left\{\sum_{i=1}^{n}|x_i-z_i+z_i-y_i|^p\right\}^{1/p} \\ &\leq \left\{\sum_{i=1}^{n}|x_i-z_i|^p\right\}^{1/p}+\left\{\sum_{i=1}^{n}|z_i-y_i|^p\right\}^{1/p} \quad \text{(using the Minkowski Inequality)} \\ &= d(x,z)+d(z,y) \end{aligned}$$

Thus,

$$d(x,y) \leq d(x,z)+d(z,y)$$

which is the triangular inequality.

**Convergent Sequences**: Consider a sequence $\{u_k\}$. We say that the sequence $\{u_k\}$ converges to $\overline{u}$, i.e.,

$$\lim_{k\to\infty} = \overline{u}$$

if for every $\epsilon > 0, \exists$ an $N$ such that

$$d(\overline{u}, u_k) \leq \epsilon \quad \forall\, k > N$$

A sequence is said to diverge if it does not coverge.

**Cauchy Sequence**: $u_k$ is said to be a Cauchy sequence if $\forall\, \epsilon > 0$, $\exists\, N$ such that,

$$d(u_i, u_j) \leq \epsilon \quad \forall i, j > N$$

Theorem: If $u_k$ converges then it is a Cauchy sequence.

Proof: If $u_k$ converges then

$$\lim_{k\to\infty} = \overline{u}$$

Using the triangular inequality

$$d(u_i, u_j) \leq d(u_i, \overline{u}) + d(\overline{u}, u_j)$$

Since $u_k$ is convergent $\exists$ an $N$ such that

$$d(u_i, \overline{u}) \leq \frac{\epsilon}{2} \qquad \text{and} \qquad d(\overline{u}, u_j) \leq \frac{\epsilon}{2}, \qquad i, j > N$$

Thus $\exists$ $N$ such that

$$d(u_i, u_j) \leq \epsilon \qquad \forall\, i, j > N$$

Note that a Cauchy sequence need not be convergent. Hence in a Cauchy sequence the distance between two points in the sequence can get arbitrarily close. However the limit value of the sequence is not mentioned and it need not exist. This issue is resolved by invoking the concept of a complete metric space.

Definition: A metric space $(X, d)$ is said to be complete if every Cauchy sequence of points from X converges to a limit in $X$.

Example: Let $X[0, 1)$ which includes the value $0$ and excludes $1$. Then the sequence $u_n = 1 - \frac{1}{n}$ is a Cauchy sequence in the space $X$ since the limit of the sequence as $n \to \infty = 1$ is excluded from the space. If $X[0, 1]$ then the sequence is convergent in $X$. Thus convergence is clearly concerned with the existence of the limit points in the underlying space.

### 6.0.1 Contraction Mapping or Fixed Point Theorem

**Contraction Mapping**: Consider the mapping $F(x)$, such that

$$x = F(x)$$

$x_0$ is a fixed point of $F$ if $x_0 = F(x_0)$. Let $(X, d)$ be a metric space and $F : X \to X$. $F(x)$ is said to be a contraction if $\exists$ a real number $k$, $0 \leq k < 1$ ($k$ independent of $x$ and $y$) such that

$$d(F(x), F(y)) \leq k\, d(x, y) \quad \forall\, x, y \in X$$

This situation is illustrated graphically in the Figure below, where the distance between two points $x$ and $y$, $d(x, y)$ is reduced upon applying the mapping $F(x)$ to each of the points.

**Theorem:** Let $(X, d)$ be a complete metric space and let $F : X \to X$ be a contraction. Then $\exists$ a unique point $x_0$ in the $X$ such that $x_0 = F(x_0)$.

**Proof:** Generate a sequence $x_n$ from the mapping $F(x)$ in the following manner,

$$x_1 \;=\; F(x)$$

$$\begin{aligned} x_2 &= F(x_1) \\ \vdots &= \\ x_n &= F(x_{n-1}) \end{aligned}$$

We first show that $x_n$ is a Cauchy sequence. Consider the distances,

$$\begin{aligned} d(x_2, x_1) &= d(F(x_1), F(x)) \leq k\, d(x_1, x) \\ d(x_3, x_2) &= d(F(x_2), F(x_1)) \leq k\, d(x_2, x_1) \leq k^2\, d(x_1, x) \\ &\vdots \\ d(x_m, x_{m-1}) &= \leq k^{m-1}\, d(x_1, x) \end{aligned}$$

Using the triangular inequality,

$$\begin{aligned} d(x_3, x_1) &\leq d(x_3, x_2) + d(x_2, x_1) \\ d(x_4, x_1) &\leq d(x_4, x_3) + d(x_3, x_1) \\ &\leq d(x_4, x_3) + d(x_3, x_2) + d(x_2, x_1) \end{aligned}$$

Generalizing, for $m > n$ and using the above results,

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \ldots,\, d(x_{n+1}, x_n) \\ &\leq k^{m-1}\, d(x_1, x) + k^{m-2}\, d(x_1, x) + \ldots,\, k^n\, d(x_1, x) \\ &= [k^{m-1}) + k^{m-2} + \ldots,\, k^n]\, d(x_1, x) \\ &\leq k^n\, [k^{m-n-1}) + k^{m-n-2} + \ldots,\, k+1] d(x_1, x) \end{aligned}$$

Since $0 \leq k < 1$,

$$d(x_m, x_n) \leq k^n \sum_{i=0}^{\infty} k^i\, d(x_1, x) = \frac{k^n}{1-k}\, d(x_1, x)$$

where we have used the summation of the geometric series,

$$\sum_{i=0}^{\infty} k^i = \frac{1}{1-k}$$

Since $0 \leq k < 1$, $d(x_m, x_n) \to 0$ as $m, n \to \infty$. Thus $x_n$ is Cauchy. Further since $(X, d)$ is a complete metric space, $x_n$ is convergent in $X$. Let

$$x_0 = \lim_{n \to \infty} x_n$$

To show that $x_0$ is a fixed point of $F(x)$ we use the continuity of the mapping $F(x)$. Since $F(x)$ is continuous,

$$x_0 = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} F(x_n) = F(\lim_{n\to\infty} x_n) = F(x_0)$$

To show that $x_0$ is unique: Assume that $x_0$ and $y_0$ are two fixed points of $F(x)$, i.e. $x_0 = F(x_0)$ and $y_0 = F(y_0)$.

$$d(x_0, y_0) = d(Fx_0, Fy_0) \leq k\, d(x_0, y_0) < d(x_0, y_0)$$

Hence $d(x_0, y_0) = 0$ and $x_0 = y_0$. Thus the fixed point is unique.

Some notes about fixed points. If $F(x_0) = x_0$ then $F^p(x_0) = x_0$. Further if $x_0$ is a fixed point of $F^p$ it need not be a fixed point of $F(x)$.

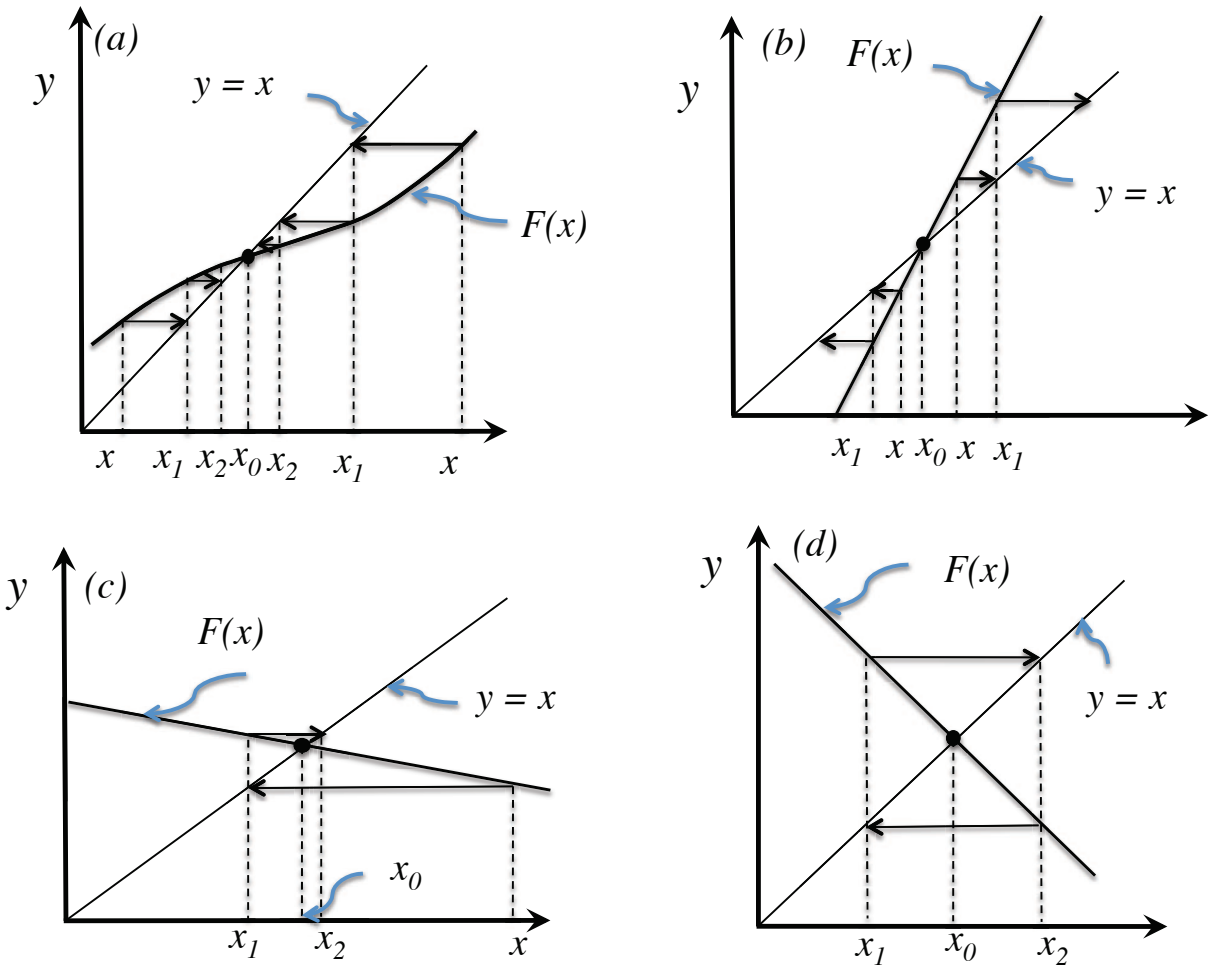



Figure 6.1: Picard iterates are illustrated for different functions $F(x)$. Convergence toward the fixed point $x_0$ is observed for cases (a) and (c). In these situations the $|F'(x)| < 1$.Since $|F'(x)| > 1$ for case (b) the iterates diverge and the iterates oscillate around the fixed point for case (d) where $|F'(x)| = 1$. In all cases the initial guess is $x$.